CHAPTER FIVE

Survey Research

CHAPTER OUTLINE

Overview

Are Americans romantic? Are they romantic compared to the French, who are renowned for their passion for passion? These were some of the questions asked in a 2009 survey of American romance—a survey conducted specifically to compare findings to a French survey regarding love and relationships (Schwartz, 2010).

Survey results indicated that Americans are just as "in love" as the French, even more so when considering older respondents. For individuals over age 65, 63% of Americans described themselves as "in love," compared to 46% of French in that age group. When do Americans and French respondents differ? When asked about sex. One question asked, "can true love exist without a radiant sex life?" A majority of Americans (77%) ages 18–65+ claimed this was true, whereas only 35% of French claimed true love can exist without such sex.

Based on these results, we can *describe* people's responses about being in love. Also, we can *predict* responses about being in love based on age and nationality (French or American). The findings also allow us to predict, knowing whether someone is American or French, what he or she may say about true love and sex. But does being French or American *cause* these attitudes? That is another matter entirely.

Key Concept

Correlational research provides a basis for making predictions. Relationships among naturally occurring variables are assessed with the goal of identifying *predictive relationships*. As we discussed in Chapter 4, a *correlation coefficient* is a quantitative index of the direction and magnitude of a predictive relationship. We will discuss correlational research in the context of survey methodology later in this chapter.

Surveys typically are conducted with samples of people. In this chapter we first introduce the basic logic and techniques of sampling—the process of selecting a subset of a population to represent the population as a whole. You will then learn about the advantages and disadvantages of various survey-research methods and survey-research designs. The primary instrument of survey research is the questionnaire, and so we describe the basics of constructing a good questionnaire. We also discuss an important question that needs to be addressed in survey research, "Do people really do what they say they do?" We conclude the chapter by critically examining a broader question, "Just what can we conclude about causality when a correlation exists between two variables?"

Uses of Surveys

- Survey research is used to assess people's thoughts, opinions, and feelings.
- Surveys can be specific and limited in scope or more global in their goals.
- The best way to determine whether results of a survey are biased is to examine the survey procedures and analyses.

We discussed in Chapter 4 how psychologists use observational methods to infer what people must have been thinking or feeling to have behaved in a certain way. Survey research is designed to deal more directly with the nature of people's thoughts, opinions, and feelings. On the surface, survey research is

deceptively simple. If you want to know what people are thinking, ask them! Similarly, if you want to know what people are doing, observe them! As we have seen, however, when we hope to infer general principles of behavior, our observations must be more sophisticated than our everyday, casual observations. So, too, survey research requires more than simply asking people questions.

Social scientists, such as political scientists, psychologists, and sociologists, use surveys in their research for a variety of reasons, both theoretical and applied. Surveys also are used to meet the more pragmatic needs of the media, political candidates, public health officials, professional organizations, and advertising and marketing directors. Surveys often are used to promote political or social agendas, as in the public health initiative to eliminate depictions of smoking in movies. Heatherton and Sargent (2009) analyzed survey data and found that as exposure to smoking in movies increases among adolescents, the likelihood of trying smoking or becoming smokers increases, especially among adolescents typically regarded as having low risk for smoking (e.g., nonsmoking parents).

In addition, the scope and purpose of surveys can be limited and specific, or they can be more global. An example of a survey with limited scope is an investigation of gratitude and communal strength in a relationship (Lambert, Clark, Durtschi, Fincham, & Graham, 2010). Communal strength refers to the degree to which individuals feel responsible for a relationship partner's welfare. Lambert and his colleagues surveyed participants to assess the extent to which individuals express gratitude in a close relationship and their feelings of communal strength in that relationship. The results of their survey supported their hypothesis that expressing gratitude is related to individuals' perception of communal strength.

Myers and Diener (1995), on the other hand, conducted a survey that addressed complex issues of global concern. They sampled people from 24 countries representing every continent but Antarctica. One of the research questions was whether people in wealthy countries have a greater sense of personal well-being than those in not-so-wealthy countries. The survey results showed that national wealth, as measured by gross national product per capita, is positively correlated with personal well-being (.67). But this relationship is not simple because national wealth is also correlated with other variables that are themselves highly correlated with well-being, such as number of continuous years of democracy (.85).

One of the ways that surveys can be used deserves mention because it raises ethical concerns. An ethical dilemma arises when sponsors of research have vested interests in the survey results. Crossen (1994) highlighted this by stating that "more and more of the information we use to buy, elect, advise, acquit, and heal has been created not to expand our knowledge but to sell a product or advance a cause" (p. 14). Crossen cites an example of a survey sponsored by a manufacturer of cellular phones showing that 70% of respondents (all of whom used cellular phones) agreed that people who use cellular telephones are more successful in business than those who do not use cell phones.

Is it reasonable to conclude that survey results are biased anytime the outcome of the survey is favorable for the sponsoring agency? Answers to ethical

questions are rarely simple, and the answer to this one is not simple. High-quality and ethical research can be done when the sponsor has an interest in the outcome. Knowing the sponsor of the research is important when evaluating survey results but is not sufficient for judging whether the study is biased. It is much more important to know whether a biased sample has been used, or whether the wording of questions has been slanted, or whether the data have been selectively analyzed or reported. Any of these aspects of survey research can bias the results, and unethical researchers can use these techniques to make the results "turn out right." The best protection against unethical researchers and poor-quality research is to examine carefully the procedures and analyses used in the survey research.

Characteristics of Surveys

- Survey research involves selecting a sample (or samples) and using a predetermined set of questions.

All properly conducted surveys share common characteristics that make surveys an excellent method for describing people's attitudes and opinions. First, surveys generally involve sampling, which is a characteristic of nearly all behavioral research. This concept was introduced in our discussion of time and situation sampling in observational research in Chapter 4. We will discuss sampling as it is used in survey research in the next section. Surveys also are characterized by their use of a set of predetermined questions for all respondents. Oral, written, or computer-entered responses to these questions constitute the principal data obtained in a survey. By using the same phrasing and ordering of questions, it is possible to summarize the views of all respondents succinctly.

When a *representative sample* of people is asked the same set of questions, we can describe the attitudes of the population from which the sample was drawn. Furthermore, when the same questions are used, we can compare the attitudes of different populations or look for changes in attitudes over time. Surveys are a powerful tool in researchers' toolbox. In the remainder of this chapter, we highlight the methods that make surveys an effective strategy for examining people's thoughts, opinions, and feelings.

Sampling in Survey Research

- Careful selection of a survey sample allows researchers to generalize findings from the sample to the population.

Assume you've decided your research question is best answered using a survey, and you've determined the population of interest for your survey. The next step is to decide who should respond to your survey questions. This involves carefully selecting a sample of respondents to represent the population. Whether describing a national population or a much smaller one (e.g., the students of one university), the procedures for obtaining a representative sample are the same.

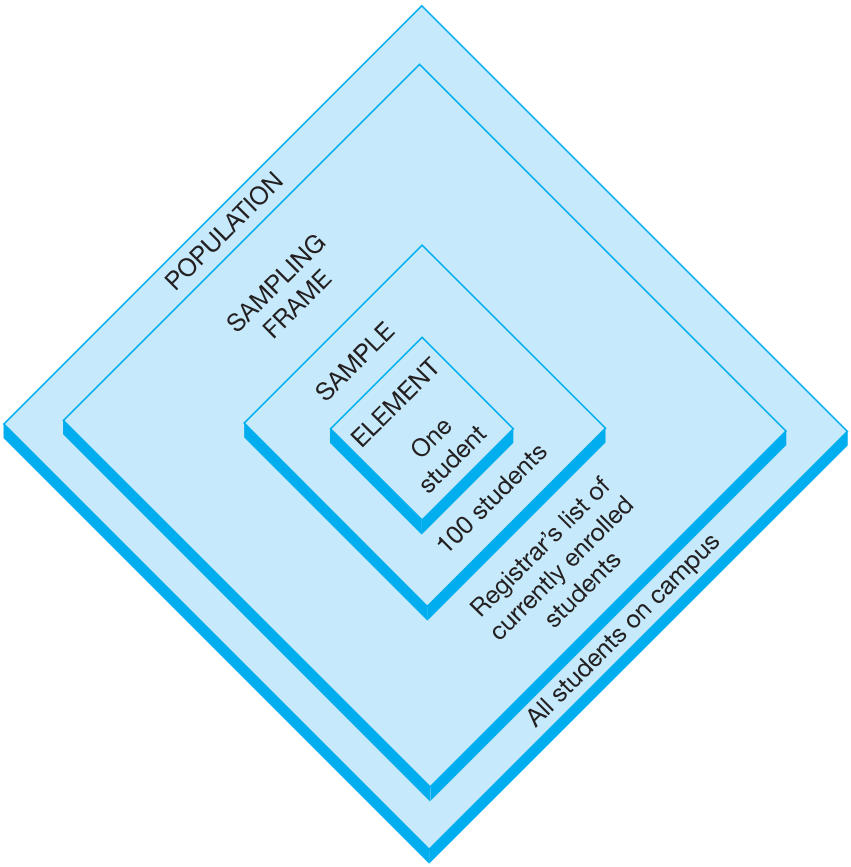
Basic Terms of Sampling

- The identification and selection of elements that will make up the sample is at the heart of all sampling techniques; the sample is chosen from the sampling frame, or list of all members of the population of interest.
- Researchers are not interested simply in the responses of those surveyed; instead, they seek to describe the larger population from which the sample was drawn.
- The ability to generalize from a sample to the population depends critically on the representativeness of the sample.
- A biased sample is one in which the characteristics of the sample are systematically different from the characteristics of the population.
- Selection bias occurs when the procedures used to select a sample result in the overrepresentation or underrepresentation of some segment(s) of the population.

Key Concept

As we begin to talk about sampling techniques, we need to be clear about the definitions of four terms: *population*, *sampling frame*, *sample*, and *element*. The relationships among the four critical sampling terms are summarized in Figure 5.1. A **population** is the set of all cases of interest. For example, if you are interested in the attitudes of students on your campus toward computer services, your population is all students on your campus. Contacting everyone in a large

FIGURE 5.1 Illustration of relationships among four basic terms in sampling.



population is often practically impossible. Therefore, researchers usually select a subset of the population to represent the population as a whole.

We need to develop a specific list of the members of the population in order to select a subset of that population. This specific list is called a *sampling frame* and is, in a sense, an operational definition of the population of interest. In a survey of students' attitudes toward computer services, the sampling frame might be a list obtained from the registrar's office of all currently enrolled students. The extent to which the sampling frame truly reflects the population of interest determines the adequacy of the sample we ultimately select. The list provided by the registrar should provide a good sampling frame, but some students might be excluded, such as students who registered late.

Key Concept

The subset of the population actually drawn from the sampling frame is called the **sample.** We might select 100 students from the registrar's list to serve as the sample for our computer survey. How closely the attitudes of this sample of students will represent all students' attitudes depends critically on how the sample is selected. Each member of the population is called an *element*. The identification and selection of elements that will make up the sample are at the heart of all sampling techniques.

It is important to emphasize at this point that samples are of little or no interest in themselves. A new computer facility is not going to be built for the sole use of the 100 students surveyed. Similarly, the social psychologist is not

STRETCHING EXERCISE I

Identifying representative samples

Presented on the left side are descriptions of four populations. Find the sample on the right side that represents each population.

Populations	Samples
1 60% women, 40% men 90% ages 18–22, 10% age >22 70% freshman/sophomore, 30% junior/senior	**A** 132 women, 44 men 114 ages 18–22, 62 age >22 141 freshman/sophomore, 35 junior/senior
2 80% women, 20% men 60% ages 18–22, 40% age >22 70% freshman/sophomore, 30% junior/senior	**B** 244 women, 61 men 183 ages 18–22, 122 age >22 213 freshman/sophomore, 92 junior/senior
3 75% women, 25% men 65% ages 18–22, 35% age >22 80% freshman/sophomore, 20% junior/senior	**C** 48 women, 12 men 54 ages 18–22, 6 age >22 42 freshman/sophomore, 18 junior/senior
4 80% women, 20% men 90% ages 18–22, 10% age >22 70% freshman/sophomore, 30% junior/senior	**D** 150 women, 100 men 225 ages 18–22, 25 age >22 175 freshman/sophomore, 75 junior/senior

From Zechmeister, Zechmeister, & Shaughnessy, *Essentials of Research Methods in Psychology,* McGraw-Hill, 2001, p. 124.

interested solely in the racial attitudes of the 50 people he surveyed, nor is the marketing director interested only in the preferences of the 200 consumers she surveyed. *Populations, not samples, are of primary interest*. The "power" of samples to describe the larger population is based on the assumption that survey responses in a sample can be applied to the population from which the sample was drawn.

Key Concept

The ability to generalize from a sample to the population depends critically on the **representativeness** of the sample. Clearly, individuals in a population differ in many ways, and populations differ from each other. For example, one population might be 40% female and 60% male, whereas in another population the distribution might be 75% female and 25% male. *A sample is representative of the population to the extent that it exhibits the same distribution of characteristics as the population*. If a representative sample of 200 adults has 80 men and 120 women, which of the above-mentioned populations does it represent? You can use the illustrations in Stretching Exercise I to gain additional practice in identifying representative samples.

The major threat to representativeness is bias. A *biased sample* is one in which the distribution of characteristics in the sample is systematically different from the target population. A sample of 100 adults that included 80 women and 20 men would likely be biased if the population were 60% female and 40% male. In this case, women would be overrepresented and men would be underrepresented in the sample. There are two sources of bias in samples: selection bias and response rate bias.

Key Concept

Selection bias occurs when the *procedures* used to select the sample result in the overrepresentation of some segment of the population or, conversely, in the exclusion or underrepresentation of a significant segment. We will describe problems associated with response rate bias in the next section, "Survey Methods."

Selection bias is likely, for example, when exit polls are used to survey people's attitudes. Research indicates that demographic characteristics such as age, race, education, and income of voters interviewed in exit polls differ from characteristics of the population based on U.S. Census data (Madigan, 1995). Note that U.S. Census data represents the entire population and includes voters and nonvoters, whereas only voters are selected for exit-poll samples. Thus, exit-poll samples may not represent the population due to a selection bias. Although a voter poll may accurately reflect the interests and attitudes of people *who vote*, their survey responses may not be used to characterize the attitudes of the population (which includes people who did not vote). Clearly, politicians cannot assume a "mandate" based on a biased sample of individuals who voted.

A more general lesson can be learned from the exit-poll example. Namely, what constitutes a representative sample depends on the population of interest. For example, if a university wants to know student drivers' opinions about on-campus parking, then the target population is college students who bring cars to campus (not college students in general). An unbiased sample would, in this case, be one that is representative of the population of students who have cars on campus.

Approaches to Sampling

- Two approaches to selecting a survey sample are nonprobability sampling and probability sampling.
- Nonprobability sampling (such as convenience sampling) does not guarantee that every element in the population has an equal chance of being included in the sample.
- Probability sampling is the method of choice for obtaining a representative sample.
- In simple random sampling, each element of the population has an equal chance of being included in the sample; in stratified random sampling, the population is divided into subpopulations (strata), and random samples are drawn from the strata.

Key Concept

There are two basic approaches to sampling—nonprobability sampling and probability sampling. In **nonprobability sampling** we have no guarantee that each element has some chance of being included and no way to estimate the probability of each element's being included in the sample. In the computer-services survey we described earlier, if a researcher interviewed the first 30 students who entered the library, she would be using nonprobability sampling. Clearly, not all students would be equally likely to be at the library at that particular time, and some students would have essentially no chance of being included in the sample (e.g., if at work or in class).

Key Concept

By contrast, if the researcher were to select 100 students randomly from the registrar's list of enrolled students, she would be using probability sampling. In **probability sampling,** all registered students (elements) have an equal chance of being included in the sample. We can describe this researcher's approach as probability sampling because her sampling procedure (i.e., random selection from a predetermined list) allows all students to have an equal chance of being selected for the survey. *Probability sampling is far superior to nonprobability sampling in ensuring that selected samples represent the population.* Thus, the researcher who selects 30 students randomly from the registrar's list of students is more likely to have a representative sample than the researcher who bases her survey results on the first 30 students who show up at the library.

Nonprobability Sampling

The most common form of nonprobability sampling is convenience sampling. *Convenience sampling* involves selecting respondents primarily on the basis of their availability and willingness to respond. For example, newspapers often publish the comments of "the person on the street." Their comments may make interesting reading, but their opinions likely do not represent those of the wider community. This lack of representativeness arises because convenience sampling is nonprobability sampling, and we can't be sure that every person in the community had a chance to be included in the sample. Convenience sampling also is involved when people respond to surveys in magazines because the magazine has to be available (and purchased), and people must be willing to send in their responses. The "participant pool" that is tapped by many psychologists at colleges and universities is a convenience

sample typically comprised of students registered for the introductory psychology course.

Crossen (1994) describes the drawbacks of another variation of convenience sampling, call-in surveys. Call-in surveys are used by TV and radio shows to poll the views of their audience. Those who happen to be "tuned in" and who are willing to call (and sometimes to pay the charge for calling a 900 number) make up the sample for these call-in surveys. People who make calls in response to a call-in request differ from the general population not only because they are part of the particular show's audience, but because they are motivated enough to make a call. Similarly, online computer users who respond to a "pop up" survey question displayed on their home page will differ from those who choose not to respond (or are not regular computer users).

A prime-time TV news show once conducted a call-in survey with a question concerning whether the United Nations (UN) headquarters should remain in the United States (Crossen, 1994). It turns out that another survey research study involving about 500 randomly selected respondents also asked the same question. Of the 186,000 callers who responded, a solid majority (67%) wanted the UN *out of the United States.* Of the 500 respondents to the survey research study, a clear majority (72%) wanted the UN *to stay in the United States.* How could these two surveys yield such different—even opposite—results? Should we put more confidence in the results of the call-in survey because of the massive sample size? Absolutely not! A large convenience sample is just as likely to be an unrepresentative sample as is any other convenience sample. As a general rule, *you should consider that convenience sampling will result in a biased sample unless you have strong evidence confirming the representativeness of the sample.*

Probability Sampling

Key Concept

The distinguishing characteristic of probability sampling is that the researcher can specify, for each element of the population, the probability that it will be included in the sample. Two common types of probability sampling are simple random sampling and stratified random sampling. Simple random sampling is the basic technique of probability sampling. The most common definition of **simple random sampling** is that every element has an equal chance of being included in the sample. The procedures for simple random sampling are outlined in Box 5.1.

One critical decision that must be made in selecting a random sample is how large it should be. For now, we will simply note that the size of a random sample needed to represent a population depends on the degree of variability in the population. For example, college students in Ivy League schools represent a more homogeneous population than college students in *all* U.S. colleges in terms of their academic abilities. At one extreme, the most homogeneous population would be one in which all members of the population are identical. A sample of one element would be representative of this population regardless of the size of the population. At the other extreme, the most heterogeneous population would be one in which each member was completely different from all other members on all characteristics. No sample, regardless of its size, could be representative of this population. Every individual would have to be included

BOX 5.1

HOW TO DRAW RANDOM SAMPLES

The following names represent a scaled-down version of a sampling frame obtained from the registrar's office of a small college campus. Procedures for drawing both a simple random sample and a stratified random sample from this list are described.

Adamski	F	Jr
Alderink	F	Sr
Baxter	M	Sr
Bowen	F	Sr
Broder	M	So
Brown	M	Jr
Bufford	M	So
Campbell	F	Fr
Carnahan	F	So
Cowan	F	Fr
Cushman	M	Sr
Dawes	M	Jr
Dennis	M	Sr
Douglas	F	Fr
Dunne	M	So
Fahey	M	Fr
Fedder	M	Fr
Foley	F	So
Gonzales	F	Jr
Harris	F	Jr
Hedlund	F	So
Johnson	F	Fr
Klaaren	F	Jr
Ludwig	M	Fr
Martinez	F	Sr
Nowaczyk	M	Jr
O'Keane	F	Sr
Osgood	M	So
Owens	F	So
Penzien	M	Jr
Powers	M	Sr
Romero	M	Fr
Sawyer	M	Jr
Shaw	M	Sr
Sonders	F	Sr
Suffolk	F	So
Taylor	F	Fr
Thompson	M	Fr
Watterson	F	Jr
Zimmerman	M	So

Drawing a simple random sample:

Step 1. Number each element in the sampling frame: Adamski would be number 1, Harris number 20, and Zimmerman number 40.

Step 2. Decide on the sample size you want to use. This is just an illustration, so we will use a sample size of 5.

Step 3. Choose a starting point in the Table of Random Numbers in the Appendix (Table A.1) (a finger stab with your eyes closed works just fine—our stab came down at column 8, row 22 at the entry 26384). Because our sampling frame ranges only from 1 to 40, we had decided *prior* to entering the table to use the left two numbers in each set of five and to go across the table from left to right. We could just as easily have decided to go up, down, or from right to left. We could also have used the middle two or the last two digits of each set of five, but one should make these decisions before entering the table.

Step 4. Identify the numbers to be included in your sampling by moving across the table. We got the numbers 26, 06, 21, 15, and 32. Notice that numbers over 40 are ignored. The same would be true if we had come across a repetition of a number we had already selected.

Step 5. List the names corresponding to the selected numbers. In our case the sample will include Nowaczyk, Brown, Hedlund, Dunne, and Romero.

An even easier system, called *systematic sampling,* can be used to obtain a random sample. In this procedure you divide the sample size you want into the size of the sampling frame to obtain the value *k*. Then you select every *k*th element after choosing the first one randomly. In our example we want a sample size of 5 from a sampling frame of 40, so *k* would be 8. Thus, we would choose one of the first eight people randomly and then take every eighth person thereafter. If Alderink were chosen from among the first eight, the remaining members of the sample would be Cowan, Foley, Nowaczyk, and Shaw. *Note:* This system should *not* be used if the sampling frame has a periodic organization—if, for example, you had a list of dormitory residents arranged by room and every 10th pair listed occupied a corner room. You can readily see that, in such a list, if your sampling interval was 10 you could end up with all people from corner rooms or no people from corner rooms.

Freshmen

1 Campbell
2 Cowan
3 Douglas
4 Fahey
5 Fedder
6 Johnson
7 Ludwig
8 Romero
9 Taylor
10 Thompson

Sophomores

1 Broder
2 Bufford
3 Carnahan
4 Dunne
5 Foley
6 Hedlund
7 Osgood
8 Owens
9 Suffolk
10 Zimmerman

Juniors

1 Adamski
2 Brown
3 Dawes
4 Gonzales
5 Harris
6 Klaaren
7 Nowaczyk
8 Penzien
9 Sawyer
10 Watterson

Seniors

1 Alderink
2 Baxter
3 Bowen
4 Cushman
5 Dennis
6 Martinez
7 O'Keane
8 Powers
9 Shaw
10 Sonders

Drawing a stratified random sample:

Step 1. Arrange the sampling frame in strata. For our example we stratified by class standing. In the example the strata are equal in size, but this need not be the case.

Step 2. Number each element within each stratum, as has been done in the foregoing list.

Step 3. Decide on the overall sample size you want to use. For our example we will draw a sample of 8.

Step 4. Draw an equal-sized sample from each stratum such that you obtain the desired overall sample size. For our example this would mean drawing 2 from each stratum.

Step 5. Follow the steps for drawing a random sample and repeat for each stratum. We used a different starting point in the Table of Random Numbers (Table A.1), but this time we used the last two digits in each set of five. The numbers identified for each stratum were Freshmen (04 and 01), Sophomores (06 and 04), Juniors (07 and 09), and Seniors (02 and 09).

Step 6. List the names corresponding to the selected numbers. Our stratified random sample would include Fahey, Campbell, Hedlund, Dunne, Nowaczyk, Sawyer, Baxter, and Shaw.

to describe such a heterogeneous population. In practice, the populations with which survey researchers work typically fall somewhere between these two extremes.

Key Concept

The representativeness of a sample can often be improved by using stratified random sampling. In **stratified random sampling,** the population is divided into subpopulations called *strata* (singular: *stratum*) and random samples are drawn from each of these strata. There are two general ways to determine how many elements should be drawn from each stratum. One way (illustrated in the last example of Box 5.1) is to draw equal-sized samples from each stratum. The second way is to draw elements for the sample on a proportional basis. Consider a population of undergraduate students made up of 30% freshmen, 30% sophomores, 20% juniors, and 20% seniors (class years are the strata). A stratified random sample of 200 students drawn from this population would include 60 freshmen, 60 sophomores, 40 juniors, and 40 seniors. In contrast, drawing equal-sized samples from each stratum would result in 50 students for each class year. *Only the stratified sample on a proportional basis would be representative.*

In addition to its potential for increasing the representativeness of samples, stratified random sampling is useful when you want to describe specific portions of the population. For example, a simple random sample of 100 students would be sufficient to survey students' attitudes on a campus of 2,000 students. Suppose, however, your sample included only 2 of the

STRETCHING EXERCISE II

Two student researchers have been asked to do a survey to determine the attitudes of students toward fraternities and sororities on campus. There are 3,200 students in the school. About 25% of the students belong to the Greek organizations and 75% do not. The two student researchers disagree about what sampling plan is best for the study. One researcher thinks they should draw a stratified random sample of 200 students: 100 from among those students who belong to Greek organizations and 100 from among the independent students. The second researcher thinks they should draw one simple random sample of 100 students from the campus as a whole.

1 Comment critically on these two sampling plans in terms of their representativeness and the likelihood that they would measure reliably the views of students who belong to Greek organizations.
2 Develop your own sampling plan if you decide that neither of the ones proposed so far is optimal.

40 chemistry majors on campus, and you wish to describe the views of students according to different majors. Although this accurately reflects the proportion of chemistry majors in the campus population, it would be risky to use the views of only 2 chemistry students to represent all 40 chemistry majors (2 is too few). In this case (and more generally when a stratum is small in number), you could sample more chemistry majors to describe their views better. We can't say precisely how many to sample because, as we learned earlier, the sample size needed to represent a population depends on the degree of variability in the population.

Survey Methods

- Four methods for obtaining survey data are mail surveys, personal interviews, telephone interviews, and Internet surveys.

Selecting the sample is only one of several important decisions to make when doing survey research. You also need to decide how you will obtain information from the respondents. There are four general methods: mail surveys, personal interviews, telephone interviews, and Internet surveys. As is often true when doing research, there is no one best survey method for all circumstances. Each survey method has its own advantages and disadvantages. The challenge you face is to select the method that best fits your research question.

Mail Surveys

- Although mail surveys are quick and convenient, there may be a problem with the response rate when individuals fail to complete and return the survey.
- Due to problems with the response rate, the final sample for a mail survey may not represent the population.

Mail surveys are used to distribute self-administered questionnaires that respondents fill out on their own. One advantage of mail surveys is that they

usually can be completed relatively quickly. Because they are self-administered, mail surveys also avoid the problems due to interviewer bias (to be defined in the next section). Among the four survey methods, mail surveys are the best for dealing with highly personal or embarrassing topics, especially when anonymity of respondents is preserved.

Unfortunately, there are many disadvantages to mail surveys. Some of these disadvantages are less serious than others. For instance, because respondents will not be able to ask questions, the questionnaire used in the survey must be completely self-explanatory. A second, less serious disadvantage is that the researcher has little control over the order in which the respondent answers the questions. The order of questions may affect how respondents answer certain questions. A serious problem with mail surveys, however, is a low response rate that can result in response rate bias.

Key Concept

Response rate refers to the percentage of people who complete the survey. For example, if 30 of 100 people sampled complete the survey, the response rate is 30%. *A low response rate indicates there could be a* ***response rate bias*** *that threatens the representativeness of a sample.* There are many reasons why this occurs. For example, respondents with literacy problems, low educational background, or vision problems may not complete the survey; therefore, people with these characteristics may not be represented well in the final sample of respondents. Often, people randomly selected for a sample are too busy or not interested enough in the study to return a completed questionnaire. Low response rate (i.e., failure to complete and return the survey) is the major factor leading to samples that do not represent the population of interest, resulting in a response rate bias. Thus, a carefully selected probability sample may become a nonprobability sample—a convenience sample in which individuals' availability and willingness determine whether they complete the survey.

Unless the return rate is 100%, the potential for response rate bias exists regardless of how carefully the initial sample was selected. However, a low response rate does not automatically indicate the sample does not represent the population. The researcher must demonstrate the extent to which the final sample of respondents who returned the survey is representative of the population, and that no segment of the population is overrepresented or underrepresented. For example, Berdahl and Moore (2006) commented that their sample likely underrepresented the harassment experiences of recent immigrants with poor English skills who may have had difficulty with the questionnaire.

The typical return rate for mail surveys is only around 30%. There are things you can do, however, to increase the return rate. Return rates generally will be higher when

—the questionnaire has a "personal touch" (e.g., respondents are addressed by name and not simply "resident" or "student");
—responding requires minimal effort from the respondent;
—the topic of the survey is of intrinsic interest to the respondent;
—the respondent identifies in some way with the organization or researcher sponsoring the survey.

Personal Interviews

- Although costly, personal interviews allow researchers to gain more control over how the survey is administered.
- Interviewer bias occurs when survey responses are recorded inaccurately or when interviewers guide individuals' responses.

When personal interviews are used to collect survey data, respondents are usually contacted in their homes or in a shopping mall, and trained interviewers administer the questionnaire. The personal interview allows greater flexibility in asking questions than does the mail survey. During an interview the respondent can obtain clarification when questions are unclear, and the trained interviewer can follow up incomplete or ambiguous answers to open-ended questions. The interviewer controls the order of questions and can ensure that all respondents complete the questions in the same order. Traditionally, the response rate to personal interviews has been higher than that for mail surveys.

The advantages of using personal interviews are impressive, but there are also a few disadvantages. Increasing fear of urban crime and an increasing number of households with no one home during the day have reduced the attractiveness of using personal interviews in the home. A significant disadvantage of conducting personal interviews is the cost. The use of trained interviewers is expensive in terms of both money and time. Perhaps the most critical disadvantage of personal interviews involves the potential for interviewer bias. The interviewer should be a neutral medium through which questions and answers are transmitted. **Interviewer bias** occurs when the interviewer records only selected portions of the respondents' answers or tries to adjust the wording of a question to "fit" the respondent. For example, suppose a respondent in a survey about television states, "The biggest problem with TV shows is too much violence." Interviewer bias would occur if the interviewer writes down "TV violence" instead of the respondent's full response. In a follow-up question, interview bias also would occur if the interviewer asked, "By violence, do you mean murders and rapes?" A more neutral probe would allow the respondent to describe what he or she means by asking, "Could you elaborate on what you mean by violence?"

Key Concept

The best protection against interviewer bias is to employ highly motivated, well-paid interviewers who are trained to follow question wording exactly, to record responses accurately, and to use follow-up questions judiciously. Interviewers should also be given a detailed list of instructions about how difficult or confusing situations are to be handled. Finally, interviewers should be closely supervised by the director of the survey project.

Computer technology makes it possible to use a hybrid of a self-administered survey and a personal interview. A person can listen to computer-recorded questions read by an interviewer and then respond to the questions on the computer. With this technology each respondent literally hears the questions read by the same interviewer in the same way, thereby reducing the risk of interviewer bias. This technology also allows respondents to answer very personal questions in relative privacy (Rasinski, Willis, Baldwin, Yeh, & Lee, 1999).

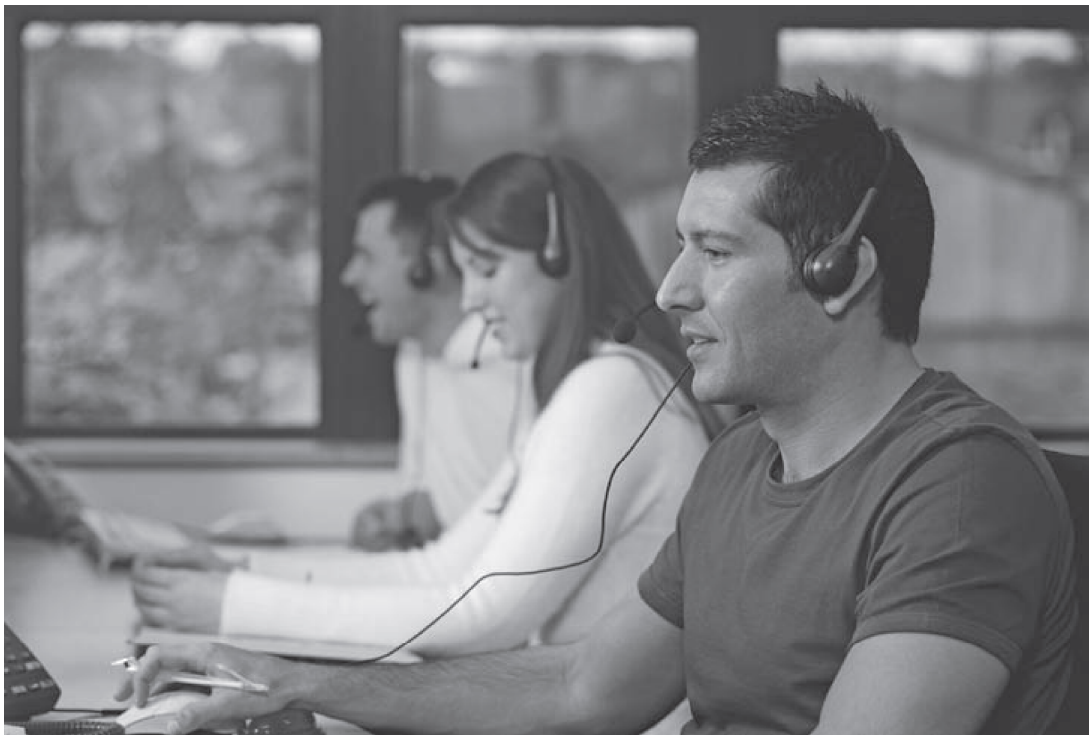
Telephone Interviews

- Despite some disadvantages, telephone interviews are used frequently for brief surveys.

The prohibitive cost of personal interviews and difficulties supervising interviewers have led survey researchers to turn to telephone or Internet surveys. Phone interviewing met with considerable criticism when it was first used because of serious limitations on the sampling frame of potential respondents. Many people had unlisted numbers, and the poor and those in rural areas were less likely to have a phone. By 2000, however, more than 97% of all U.S. households had telephones (U.S. Census Bureau, 2000), and households with unlisted numbers could be reached using random-digit dialing. The random-digit dialing technique permits researchers to contact efficiently a generally representative sample of U.S. telephone owners. Telephone interviewing also provides better access to dangerous neighborhoods, locked buildings, and respondents available only during evening hours (have you ever been asked to complete a telephone survey during dinner?). Interviews can be completed more quickly when contacts are made by phone, and interviewers can be better supervised when all interviews are conducted from one location (Figure 5.2).

The telephone survey, like the other survey methods, is not without its drawbacks. A possible selection bias exists when respondents are limited to those who have telephones and the problem of interviewer bias remains. There is a limit to how long respondents are willing to stay on the phone, and individuals

FIGURE 5.2 Random-digit dialing allows researchers efficient access to a generally representative sample of telephone owners for brief surveys.



may respond differently when talking to a "faceless voice" than they would to a personal interviewer. The proliferation of cell phones also adds an unknown effect, given that cell phone users are frequently "on the go" or in business settings when they answer their phone. This cultural change may result in lower response rates from telephone surveys. In addition, one may assume that individuals from higher socioeconomic groups are more likely to have multiple phone numbers and hence might be overrepresented in a survey based on random-digit dialing. Hippler and Schwarz (1987) suggest that people take less time to form judgments during phone interviews and may have difficulty remembering the response options offered by the interviewer. Moreover, extensive use of phone solicitation for selling products and requesting contributions has led many people to be less willing to be interviewed. Options that allow for screening calls and voice mail have made it easier for people to avoid unwanted calls. And many people who are working two jobs are rarely at home to answer the phone. In spite of these limitations and perhaps others you can think of, telephone interviews are frequently used for brief surveys.

Internet Surveys

- The Internet offers several advantages for survey research because it is an efficient, low-cost method for obtaining survey responses from large, potentially diverse and underrepresented samples.
- Disadvantages associated with Internet survey research include the potential for response rate bias and selection bias, and lack of control over the research environment.

Surveys were among the earliest Internet-based behavioral studies. Participants complete a questionnaire online and click on a "submit" button to have their responses recorded. Depending on the sophistication of the software, there is the potential for literally millions of responses to be automatically recorded and summarized as they are processed by the receiving server. Programs also exist to permit manipulation of variables and the random assignment of participants to experimental conditions. (See, for example, Fraley, 2004, for a "beginner's guide" to HTML-based psychological research on the Internet, and Kraut et al., 2004, for useful Internet resources.)

Numerous advantages of using the Internet for survey research immediately come to mind. At the top of the list are efficiency and cost (e.g., see Buchanan, 2000; Skitka & Sargis, 2005). Thousands, if not millions, of participants who vary in age, ethnicity, and even nationality can be contacted through a few keystrokes on a computer. Time and labor are dramatically reduced relative to mail or telephone surveys, let alone personal interviews. Online questionnaires are paperless, thus saving natural resources and copying costs. Participants may respond when it is convenient and do so without leaving the comfort of their home, office, dorm room, or other Internet site.

In addition to reaching large and potentially diverse samples, Skitka and Sargis (2005) suggest that the Internet also has the potential for accessing groups that typically are underrepresented in psychological research. The prevalence on the Web of chat rooms, special interest groups, and support groups provides an

"in" for a researcher seeking specific samples of participants, whether it be pet owners, members of hate groups, cancer survivors, victims of various crimes, or any of a multitude of respondent types that may not be as easily reached by traditional survey methods. Because the Internet is truly a worldwide source of participants, it also opens up new possibilities for cross-cultural research (e.g., Gosling et al., 2004).

Internet-based surveys are also not without their disadvantages. At the top of this list is the potential for sample biases (Birnbaum, 2000; Kraut et al., 2004; Schmidt, 1997). Both response rate bias and selection bias are likely to be present. Problems with low response rates can occur due to nonresponding just as it does for other survey methods. In fact, response rates typically are lower for online surveys than for comparable mail or telephone surveys (see Kraut et al., 2004; Skitka & Sargis, 2005). As we have seen, individuals who respond to a survey are going to differ on important characteristics from those who do not respond. Selection bias is present because respondents are a convenience sample comprised of individuals who have Internet access. Higher income households in the United States are more likely to have Internet access, and those households with children are more likely to have access than those without children. White and Asian householders are nearly twice as likely to have Internet access as those householders who are Black or Hispanic (Newburger, 2001).

Selection biases can be exaggerated due to the method of soliciting participants. Researchers can obtain samples of respondents by posting research notices on websites that promote research opportunities (e.g., the website associated with APS identified in Chapter 1) or by simply creating a Web page with the survey (e.g., *personality survey*) and wait for users to locate it ("hits") via Internet search engines (Krantz & Dalal, 2000). More active strategies include sending notices of the research project to individuals or groups likely to respond because of their interest in the survey topic. As Skitka and Sargis (2005) emphasize, however, not only are Internet users not representative of the general population, but also members of Internet special interest groups are not necessarily representative of their specific groups. At present there is no way to generate a random sample of Internet users (Kraut et al., 2004).

Lack of control over the research environment is also a major disadvantage of Internet surveys (Birnbaum, 2000; Kraut et al., 2004). As we mentioned in Chapter 3, this lack of control raises serious ethical issues related to informed consent and protecting individuals from harm as a consequence of their participation (e.g., emotional distress over survey questions). Because the researcher is not present, there is no easy way to determine if respondents have a clear understanding of the instructions, are answering conscientiously and not frivolously or even maliciously, or are creating multiple submissions (e.g., Kraut et al., 2004). Respondents may participate alone or in groups, under distracting conditions, without the knowledge of the researcher (Skitka & Sargis, 2005). One Internet researcher worried that respondents to survey questions about probability and risk were using calculators even though instructions requested them not to (Birnbaum, 2000). It seems safe to say that the advantages of Internet surveys outweigh many of the disadvantages. As technology improves and

IRB committees devise acceptable methods for protecting human participants, survey research on the Internet will continue to improve as a method for collecting survey data.

Survey-Research Designs

- The three types of survey design are the cross-sectional design, the successive independent samples design, and the longitudinal design.

One of the most important decisions survey researchers must make is the choice of a research design. A survey-research design is the overall plan or structure used to conduct the entire study. There are three general types of survey-research designs: the cross-sectional design, the successive independent samples design, and the longitudinal design. There is no all-purpose survey-research design. Researchers choose a design based on the goals of the study.

Cross-Sectional Design

- In the cross-sectional design, one or more samples are drawn from the population(s) at one time.
- Cross-sectional designs allow researchers to describe the characteristics of a population or the differences between two or more populations, and correlational findings from cross-sectional designs allow researchers to make predictions.

Key Concept

The cross-sectional design is one of the most commonly used survey-research designs. In a **cross-sectional design,** one or more samples are drawn from the population *at one time*. The focus in a cross-sectional design is description—describing the characteristics of a population or the differences among two or more populations at a particular time. For example, a cross-sectional design was used in a nationwide study of Internet use among 1,100 teens aged 12–17 (Lenhart, Madden, & Hitlin, 2005). Using random-digit dialing, they conducted a telephone survey of parents and teens as part of the Pew Internet and American Life Project, which is designed to examine the impact of the Internet on children, families, communities, the workplace, schools, health care, and civic/political life.

Although their findings are too numerous to describe fully here, Lenhart and her colleagues presented data that give a detailed description of teens' use of the Internet and other technology. For example, close to 9 in 10 teens reported using the Internet (compared to 66% of adults), and half of the teens reported being online at least daily. In addition, 81% of teens play games online, 76% get news online, 42% have made purchases online, and 31% reported using the Internet to get health information. Although e-mail was popular, instant messaging (IM) was preferred. Approximately 75% of the online teens in their survey (compared to 42% of online adults) use instant messaging, with half of these teens using IM every day. In fact, teens commented that they view e-mail as something for talking to "old people," institutions, or large groups.

These researchers also examined relationships among demographic variables and Internet-use variables. For example, Lenhart et al. (2005) noted that teens who are online are more likely to live in families with higher income and greater access to technology, and are disproportionately likely to be White or English-speaking Hispanic teens.

Cross-sectional designs are ideally suited for the descriptive and predictive goals of survey research. Surveys are also used to assess changes in attitudes or behaviors over time and to determine the effect of some naturally occurring event, such as the effect of the economic collapse of 2008. For these purposes the cross-sectional design is not the method of choice. Rather, research designs are needed that systematically sample respondents over time. Two such designs are discussed in the next two sections.

Successive Independent Samples Design

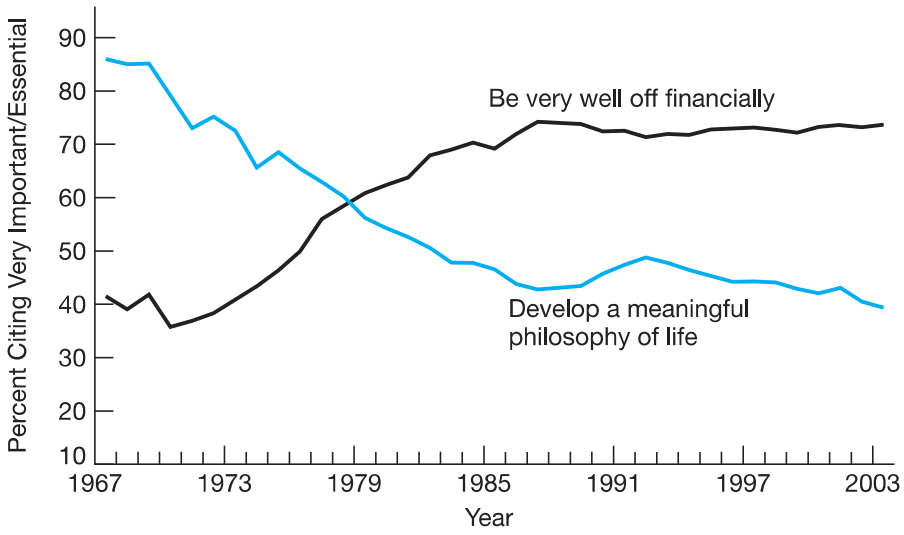
- In the successive independent samples design, different samples of respondents from the population complete the survey over a time period.
- The successive independent samples design allows researchers to study changes in a population over time.
- The successive independent samples design does not allow researchers to infer how individual respondents have changed over time.
- A problem with the successive independent samples design occurs when the samples drawn from the population are not comparable—that is, not equally representative of the population.

Key Concept

In the **successive independent samples design,** a series of cross-sectional surveys are conducted over time (successively). The samples are independent because a *different* sample of respondents completes the survey at each point in time. There are two key ingredients: (1) The same set of questions should be asked of each sample of respondents, and (2) the different samples should be drawn from the same population. If these two conditions are met, researchers can legitimately compare survey responses over time. This design is most appropriate when the major goal of the study is to describe changes in the attitudes or behaviors within a population over time. For example, public opinion researchers frequently ask independent samples of Americans the extent to which they approve of the U.S. president (referred to as the president's "approval ratings"). Changes in approval ratings over time are used to characterize Americans' opinions of the president's actions.

As another example, consider a study that you may have been part of, one that has been conducted every year since 1966. Each year some 350,000 full-time freshmen from a nationally representative sample of approximately 700 colleges and universities are surveyed (Pryor, Hurtado, DeAngelo, Patuki Blake, & Tran, 2009; Sax et al., 2003). This research project represents the largest and longest empirical study of higher education in the United States, with over 1,500 universities and over 10 million students participating over the 40-plus years of the study. Students are asked approximately 40 questions covering a number of topics, and although some changes have

FIGURE 5.3 Contrasting trends in values for college freshmen from 1966 to 2003.
Source: Sax et al. (2003), Figure 7 (p. 7).



occurred in the questions over the decades, many questions have been asked each year, making this an excellent example of a successive independent samples design.

What can be said about changes in students' values and goals during this time period? Sax et al. (2003) reported the results for the portion of the survey in which students are asked to rate the importance of different values to assess students' need for meaning and purpose in their life. Two values were of particular interest: "the importance of developing a meaningful philosophy of life" and "the importance of being very well off financially" (pp. 6–7). Figure 5.3 displays the results for the percentage of students who endorsed these values as "very important" or "essential." In the late 1960s, over 80% of students indicated that developing a meaningful philosophy of life was very important or essential—in fact, this was the top value endorsed by students. In contrast, being well-off financially was very important or essential to less than 45% of the students, and ranked fifth or sixth among students' values during the late 1960s.

In 2003, the position of these values was reversed, with 73.8% of students endorsing being well-off financially as very important or essential. In 2003, developing a meaningful philosophy of life dropped to its lowest value in the survey history, with 39.3% of students endorsing this as very important or essential. As can be seen in Figure 5.3, these contrasting trends in values began to shift in the early 1970s, crossed in 1977, and were completely reversed by the late 1980s. Sax et al. (2003) emphasize that the contrasting trends in values since the late 1980s "reflect the continuing tension between extrinsic and intrinsic values within this generation of college students" (p. 7). Data from the 2009 sample may be used to illustrate the effect of a natural treatment—the dramatic collapse of the world economy near the end of 2008. In the 2009 sample, a record 78.1% of freshmen identified "being

well-off financially" as a very important or essential objective, higher than any other item on the survey (Pryor et al., 2009). The researchers cited the economic downturn as an important factor in students' responses to the survey, including items reflecting the increasing financial difficulties associated with attending college.

The successive independent samples design has limitations. Consider hypothetical results from a successive independent samples design. Suppose you hear it reported that in 1977, 35% of college students surveyed said they don't trust the U.S. government, 25% reported they have mixed feelings, and 40% reported they do trust the U.S. government. Then you hear it reported that in 2007 results to the same survey question showed that 55% of students say they don't trust the government, 25% say they have mixed feelings, and 20% do have trust. How can we interpret these results? To account for the attitude change in the 2007 sample, can we conclude, for example, that 20% of the 1977 "do trust" group changed their minds and now don't trust the government? No! And perhaps you can see why.

What we must remember is that the students surveyed in 1977 (in our hypothetical survey) were not the same students surveyed in 2007. The extent to which specific individuals change their views over time can be determined only by testing the *same* individuals on both occasions. We cannot determine in the successive independent samples design who has changed their views or by how much. You may have considered a similar problem of interpretation when examining the results of the Sax et al. (2003) survey presented in Figure 5.3. What accounts for the changes in students' attitudes observed from 1966 to 2003? We can't say on the basis of these data. The purpose of the successive independent samples design is to describe changes over time in the distribution of *population* characteristics, not to describe changes in *individual* respondents. Accordingly, the successive independent samples design is not always helpful in ferreting out reasons for observed changes like those shown in Figure 5.3. (As you will soon see, another survey design, the longitudinal design, is more appropriate in these situations.)

A second potential limitation of the successive independent samples design arises when the successive samples are not representative of the same population. Imagine that in our hypothetical survey of students' attitude toward the U.S. government, the sample comprised students from small rural colleges in 1977 and students from large urban universities in 2007. The comparisons of students' attitudes toward the government over this time period would be meaningless. That is, we wouldn't be able to state that the student population had become less trusting over time because it's possible that the degree of trust differs for rural and urban students, which could also account for the difference between 1977 and 2007 results. The rural and urban samples illustrate the problem of *noncomparable successive samples. Changes in the population across time can be described accurately only when the successive independent samples represent the same population*. Although sophisticated statistical procedures exist to help unravel the problems associated with noncomparable successive samples, the best solution is to avoid the problem by carefully selecting successive samples that represent *the same* population.

Longitudinal Design

- In the longitudinal design, the same respondents are surveyed over time in order to examine changes in individual respondents.
- Because of the correlational nature of survey data, it is difficult to identify the causes of individuals' changes over time.
- As people drop out of the study over time (attrition), the final sample may no longer be comparable to the original sample or represent the population.

Key Concept

The distinguishing characteristic of the **longitudinal design** is that the same sample of respondents is surveyed more than once. The longitudinal design has two important advantages. First, the investigator can determine the direction and extent of change for individual respondents. Also, because changes in each individual's responses are assessed, it's easier to investigate reasons for attitude or behavior changes. Second, the longitudinal design is the best survey design when a researcher wishes to assess the effect of some naturally occurring event.

For example, Lucas (2005) examined changes in life satisfaction before and after divorce in an 18-year longitudinal study of German households that began in 1984. Many cross-sectional surveys have demonstrated that divorced people are less satisfied with life than are married people. Lucas sought to determine if divorce causes lower life satisfaction. Results indicated that these individuals' life satisfaction dropped before the divorce and gradually increased again following the divorce but did not return to their baseline state, indicating that the divorce likely decreased life satisfaction. However, Lucas also discovered that people who eventually divorced were less satisfied at the beginning of the study than those who stayed married—even before either group was married. Lucas concluded that the relationship between divorce and life satisfaction is due to preexisting differences in life satisfaction and to lasting changes due to divorce.

Heatherton, Keel, and their colleagues have used the longitudinal design to investigate changes in attitudes and behaviors related to eating during the transitions from college to early adulthood and from early adulthood to middle years (Heatherton, Mahamedi, Striepe, Field, & Keel, 1997; Keel, Baxter, Heatherton, & Joiner, 2007). Although much is known about eating disorders in adolescents and college students, less information is available about how disordered eating may progress as individuals settle down, marry, establish careers, raise children, and gain a stronger sense of identity. These researchers hypothesized that as individuals change their roles and life goals during adulthood, their emphasis on physical appearance may decrease, which would decrease the prevalence of eating disordered attitudes and behaviors (see Figure 5.4).

The first "panel" of the study took place in 1982, when a randomly selected sample of 800 women and 400 men from a private northeastern college was asked to complete a survey about eating and dieting. The response rate was 78% ($N = 625$) for women and 69% ($N = 276$) for men. In 1992 the researchers contacted these same individuals (with the help of the alumni office) and gave them the same survey again about their eating attitudes and behaviors. The third panel of data was collected in 2002, when the same individuals were in their early forties. The distinguishing characteristic of the longitudinal design is

FIGURE 5.4 Survey research such as that of Heatherton, Keel, and their colleagues (1997; 2007) investigates how individuals are affected by eating disorders as they grow older.

the fact that the *same* individuals were surveyed in each phase of the study. Although longitudinal designs involve a massive effort, the potential power of such an effort is that researchers can examine changes within individuals over time.

The researchers observed that eating attitudes and behaviors changed over time. In the decade after college, women's eating-disorder symptoms, chronic dieting, and body dissatisfaction decreased (Heatherton et al., 1997). However, despite these decreases, women's dissatisfaction with their body and their desire to lose weight remained high. Men, in contrast, rarely had problems with eating and weight during college. Ten years later, however, they had experienced weight gain (an average of almost 12 pounds, compared to women's average gain of 4 pounds). Men also reported increased dieting and symptoms of disordered eating in the 10 years after college, although this was still low relative to women.

Heatherton et al. (1997) made some interesting observations that are relevant to our understanding of longitudinal surveys. They proposed that decreases in women's eating problems reflect their maturation during their 20s, changes in their roles, and being away from the college campus (and the pressures to be thin that occur on college campuses). It's possible, however, that other processes may account for changes within the individuals in the sample. Using a successive independent samples design in which *separate* samples of college students were surveyed in 1982 and 1992, Heatherton, Nichols, Mahamedi, and Keel (1995) noted that eating-disorder symptoms and body dissatisfaction also were lower for the college students in the 1992 sample relative to the 1982

sample. These findings suggest that decreases in eating-disorder attitudes and behaviors may reflect changes at a societal level over the 10-year period (e.g., due to increasing information about eating disorders in the media). One potential problem with longitudinal survey designs is that it is difficult to pinpoint the exact causes for individuals' changes over time.[1]

What can be said about eating attitudes and behaviors 20 years following college? Overall, women demonstrated more weight dissatisfaction, dieting, and eating-disorder attitudes than men across the 20 years of the survey (Keel et al., 2007). In the 2002 survey, researchers observed that, on average, body weight increased significantly for both men (17 pounds since college) and women (14 pounds since college). Men's dieting and weight dissatisfaction was greatest in 2002, paralleling their weight gain. Interestingly, by the time the women in the study were in their early forties, despite their weight gain, they reported less dieting, less disordered eating, and less dissatisfaction with their body. In fact, women's greatest dissatisfaction with their body occurred while in college. Based on their statistical analyses, Keel et al. suggested that adult roles attained through marriage, parenthood, and careers were associated with decreases in women's disordered eating. That is, while physical appearance was important during college years (e.g., for attracting a potential mate), changes in priorities associated with marriage and becoming a mother made women's desire for thinness less important.

Another potential problem with longitudinal designs is that it can be difficult to obtain a sample of respondents who will agree to participate over time in a longitudinal study. In addition, you might think the longitudinal design solves the problem of noncomparable samples because the same people participate over and over (so of course the sample represents the same population each time). Unfortunately, the samples over time in a longitudinal design are identical *only if* all members of the original sample participate throughout the study. This is unlikely. For example, in the Heatherton et al. (1997) study, of the 901 participants in the original 1982 sample, only 724 (80%) returned a usable survey in 1992. In the third panel in 2002, 654 (73%) of the original 900 participants from 1982 responded to the survey and of these, 561 (86%) also responded to the 1992 survey. Thus, by the end of the 20 years, the researchers had survey responses for each of the three time periods (1982, 1992, 2002) for 62.3% of their original sample of 900 respondents.

Unless all the respondents in the original sample complete all phases of a longitudinal design, there is a possible problem due to *attrition*. Attrition is probably the most serious disadvantage of the longitudinal design because as samples decrease over time, they are less likely to represent the original population from which the sample was drawn. It is usually possible, however, to determine whether the final sample is comparable to the original sample in a longitudinal design. The characteristics of nonrespondents in the follow-up phase(s) are known because they participated in the original sample. Therefore, researchers

[1]Heatherton et al. (1997) noted that because the decreases in problem eating were larger among individuals in the longitudinal survey than in the successive independent samples survey, maturational processes within individuals, in addition to societal changes, likely were operating to decrease problem eating over time.

can look at characteristics of original participants to see how these nonresponding individuals may differ from those who continued their participation.

Keel et al. (2007) examined problems associated with attrition by comparing the responses of individuals who responded to the original 1982 survey but did not continue (nonrespondents) to responses of individuals who continued the study through the 2002 survey. They found that, compared to nonrespondents, individuals who continued to participate in the study described themselves as heavier, dieting more frequently, and had a greater desire for thinness. This represents a potential response rate bias because continued participation in 2002 may have been related to interest in the survey topic. Keel et al. suggested that weight and body concerns in the 2002 survey may have been inflated because of this potential response rate bias.

The advantages of the longitudinal design, such as determining changes for individual respondents, arise because the same individuals are surveyed more than once. Paradoxically, problems can also arise in longitudinal designs because of this same feature. One possible problem is that respondents may strive heroically to be consistent across surveys. This can be particularly troublesome if the study is designed to assess changes in respondents' attitudes! Although their attitudes have actually changed, people may report their original attitudes in an effort to appear consistent (perhaps they know researchers value reliability). Another potential problem is that the initial survey may sensitize respondents to the issue under investigation. For example, consider a longitudinal design used to assess students' concern about crime on campus. Once the study starts, participants may pay more attention to crime reports than they normally would. You might recognize this as an illustration of reactive measurement—people behaving differently because they know they are participating in a study.

Rather than trying to be heroically consistent in their eating attitudes and behaviors over time, Heatherton et al. (1997) noted that their participants may have been reluctant to report that they were having the same problems with eating as when they were in college. Thus, the decreases the researchers observed in problem eating during the 10-year period may be due to the fact that "women who are approaching their thirties may be embarrassed to admit they are experiencing problems typically associated with adolescence" (p. 124). When survey respondents are asked to report their attitudes and behaviors, researchers must be alert to reasons why their respondents' reports may not correspond to their actual behavior. We will return to this important issue later in this chapter.

QUESTIONNAIRES

Even if the sample of respondents was perfectly representative, the response rate was 100%, and the research design was elegantly planned and perfectly executed, the results of a survey will be useless if the questionnaire is poorly constructed. In this section we describe the most common survey research instrument, the questionnaire. To be useful, questionnaires should yield reliable and valid measures of demographic variables and of individual differences on

self-report scales. Although there is no substitute for experience when it comes to preparing a good questionnaire, there are a few general principles of questionnaire construction with which you should be familiar. We describe six basic steps in preparing a questionnaire and then offer specific guidelines for writing and administering individual questions.

Questionnaires as Instruments

- Most survey research relies on the use of questionnaires to measure variables.
- Demographic variables describe the characteristics of people who are surveyed.
- The accuracy and precision of questionnaires requires expertise and care in their construction.
- Self-report scales are used to assess people's preferences or attitudes.

Key Concept

The value of survey research (and any research) ultimately depends on the quality of the measurements that researchers make. The quality of these measurements, in turn, depends on the quality of the instruments used to make the measurements. The primary research instrument in survey research is the **questionnaire.** On the surface, a questionnaire may not look like the high-tech instruments used in much modern scientific research; but, when constructed and used properly, a questionnaire is a powerful scientific instrument for measuring different variables.

Demographic Variables Demographic variables are an important type of variable frequently measured in survey research. Demographic variables are used to describe the characteristics of the people who are surveyed. Measures such as race, ethnicity, age, and socioeconomic status are examples of demographic variables. Whether we decide to measure these variables depends on the goals of our study, as well as on other considerations. For example, Entwisle and Astone (1994) noted that "the ethnic and racial diversity of the U.S. population is now projected to increase through the middle of this 21st century, so that by then the majority of the U.S. population will be persons whose ethnicity would now be classified as 'nonwhite'" (p. 1522). By asking respondents to identify their race and ethnicity, we are able to document the mix of our sample and, if related to our research questions, compare groups according to race and ethnicity.

Measuring a demographic variable such as race may at first seem very easy. One straightforward method is simply to ask respondents to identify their race in an open-ended question: What is your race? _____ Such an approach may be straightforward, but the resulting measurement of race may not be satisfactory. For example, some respondents may mistakenly confuse "race" and "ethnicity." Important distinctions in identifying ethnic groups may go unrecognized by respondents and researchers. For instance, Hispanic does not identify a race; Hispanic designates all those whose country of origin is Spanish speaking. So, a person born in Spain would be classified as Hispanic. Latino is a term that is sometimes used interchangeably with Hispanic, but Latino designates people whose origin is from the countries of North and

FIGURE 5.5 Although ethnic background is an important demographic variable, accurately classifying people on this variable is not an easy task.

South America, excluding Canada and the United States. Distinctions like these can be confused (see Figure 5.5). For example, a person known to the authors is of European Spanish heritage and correctly considers himself a Caucasian, and not Latino. His ethnicity is Hispanic.

In general, "quick and dirty" approaches to measurement in survey research tend to yield messy data that are hard to analyze and interpret. For example, many individuals identify themselves as "multi-racial"; however, if researchers fail to include this as a possible response option, the information from participants may be incorrect—or they may skip the question entirely. Entwisle and Astone (1994) recommend a deliberate—and effective—approach when measuring race. They outline a series of nine questions to measure a person's race. One of these questions is "What race do you consider yourself to be?" Other questions seek information such as what countries the person's ancestors came from and whether Latino respondents are Mexican, Puerto Rican, Cuban, or something else. This more detailed series of questions allows researchers to measure race and ethnicity less ambiguously, more accurately, and more precisely. We use this example of measuring race and ethnicity to illustrate a more general principle: *The accuracy and precision of questionnaires as survey-research instruments depends upon the expertise and care that go into their construction.*

Preferences and Attitudes Individuals' preferences and attitudes are frequently assessed in surveys. For example, a marketing researcher may be interested in consumers' preferences for different brands of coffee, or a political group may be interested in potential voters' attitudes regarding controversial public issues. Psychologists have long been interested in measuring people's thoughts and

feelings on a vast array of topics, and often develop self-report scales for people to provide oral or written responses to items on the scale.

Self-report scales are commonly used to measure people's judgments about items presented on the scale (e.g., divorce, political candidates, life events) or to determine differences among people on some dimension presented on the scale (e.g., personality traits, amount of stress). For example, respondents may be asked to rate different life events according to how stressful they perceive the events to be. The researcher then may develop a list of life events that vary on the dimension of stressfulness. This type of scale focuses on differences among the items on the scale, not differences among individuals. To measure individual differences, respondents may be asked to report how often during the past year they experienced different stressful life events listed on a scale. A total stress score can be obtained for each individual by summing responses to the items on the scale. Individuals can then be compared according to the amount of stress experienced during the past year.[2]

Self-report measures, often in the form of a questionnaire, are among the most frequently used tools in psychology. Given their importance, it is critical that these measures be developed carefully. Two critical characteristics of the measurements made using self-report questionnaires are essential characteristics of all measurements—reliability and validity.

Reliability and Validity of Self-Report Measures

- Reliability refers to the consistency of measurement and is frequently assessed using the test–retest reliability method.
- Reliability is increased by including many similar items on a measure, by testing a diverse sample of individuals, and by using uniform testing procedures.
- Validity refers to the truthfulness of a measure: Does it measure what it intends to measure?
- Construct validity represents the extent to which a measure assesses the theoretical construct it is designed to assess; construct validity is determined by assessing convergent validity and discriminant validity.

Reliable self-report measures, like reliable observers or any other reliable measurements, are characterized by consistency. A reliable self-report measure is one that yields similar (consistent) results each time it is administered. Self-report measures must be reliable when making predictions about behavior. For example, in order to predict stress-related health problems, measures of individuals' life stress must be reliable. There are several ways to determine a test's reliability. One common method is to compute a *test–retest reliability*. Usually, test–retest reliability involves administering the same questionnaire to a large sample of people at two different times (hence, test and retest). For a questionnaire to yield reliable measurements, people need not obtain identical scores on

[2]The area of psychological measurement concerned with scaling items or stimuli is known as psychophysics, and the area of measurement concerned with individual differences is referred to as psychometrics.

the two administrations of the questionnaire, but a person's relative position in the distribution of scores should be similar at the two test times. The consistency of this relative positioning is determined by computing a correlation coefficient using the two scores on the questionnaire for each person in the sample. A desirable value for test–retest reliability coefficients is .80 or above, but the size of the coefficient will depend on factors such as the number and types of items.

A self-report measure with many items to measure a construct will be more reliable than a measure with few items. For example, we are likely to have unreliable measures if we try to measure a baseball player's hitting ability based on a single time at bat or a person's attitude toward the death penalty based on a single question on a survey. The reliability of our measures will increase greatly if we average the behavior in question across a large number of observations—many at-bats and many survey questions (Epstein, 1979). Of course, researchers must walk a fine line between too few items and too many items. Too many items on a survey can cause respondents to become tired or careless about their responses.

In general, measurements will also be more reliable when there is greater variability on the factor being measured among the individuals being tested. Often the goal of measurement is to determine the extent to which individuals differ. A sample of individuals who vary a great deal from one another is easier to differentiate reliably than are individuals who differ by only a small amount. Consider this example. Suppose we wish to assess soccer players' ability to pass the ball effectively to other players. We will be able to differentiate more reliably good players from poor players if we include in our sample a wider range of players—for example, professionals, high school players, and peewee players. It would be much harder to differentiate players reliably if we tested only professional players—they'd all be good! Thus, a test is often more reliable when administered to a diverse sample than when given to a restricted sample of individuals.

A third and final factor affecting reliability is related to the conditions under which the questionnaire is administered. Questionnaires will yield more reliable measurements when the testing situation is free of distractions and when clear instructions are provided for completing the questionnaire. You may remember times when your own test performance was hindered by noise or when you weren't sure what a question was asking.

The reliability of a survey measure is easier to determine and to achieve than the validity of a measure. The definition of validity is deceptively straightforward—a valid questionnaire measures what it is intended to measure. Have you ever heard students complain that questions on a test didn't seem to address the material covered in class? This is an issue of validity.

At this point, we will focus on construct validity, which is just one of the many ways in which the validity of a measurement is assessed. The *construct validity* of a measure represents the extent to which it measures the theoretical construct it is designed to measure. One approach to determining the construct validity of a test relies on two other kinds of validity: convergent validity and discriminant validity. These concepts can best be understood by considering an example.

TABLE 5.1 EXAMPLE OF CONSTRUCT VALIDITY*

	SWLS	LS-5	PA
SWLS	(.88)		
LS-5	.77	(.90)	
PA	.42	.47	(.81)

*Data from Lucas et al. (1996), Table 3.
Note: SWLS = Satisfaction with Life Scale; LS-5 = 5-item Life Satisfaction scale; PA = Positive Affect scale.

Table 5.1 presents data showing how we might assess the construct validity of a measure of "life satisfaction." Lucas, Diener, and Suh (1996) note that psychologists are increasingly examining factors such as happiness, life satisfaction, self-esteem, optimism, and other indicators of well-being. However, it's not clear whether these different indicators all measure the same construct (e.g., well-being) or whether each is a distinguishable construct. Lucas and his colleagues conducted several studies in which they asked individuals to complete questionnaire measures of these different indicators of well-being. For our purposes we will focus on a portion of their data from their third study, in which they asked participants to complete three scales: two life satisfaction measures, the Satisfaction with Life Scale (SWLS) and a 5-item Life Satisfaction measure (LS-5); and a measure of Positive Affect (PA). At issue in this example is whether the construct of life satisfaction—the quality of being happy with one's life—can be distinguished from being happy more generally (positive affect).

The data in Table 5.1 are presented in the form of a correlation matrix. A correlation matrix is an easy way to present a number of correlations. Look first at the values in parentheses that appear on the diagonal. These parenthesized correlation coefficients represent the values for the reliability of each of the three measures. As you can see, the three measures show good reliability (each is above .80). Our focus, however, is on measuring the construct validity of "life satisfaction," so let's look at what else is in Table 5.1.

It is reasonable to expect that scores on the Satisfaction with Life Scale (SWLS) should correlate with scores on the 5-item Life Satisfaction measure; after all, both measures were designed to assess the life satisfaction construct. In fact, Lucas et al. observed a correlation between these two measures of .77, which indicates that they correlate as expected. This finding provides evidence for *convergent validity* of the measures; the two measures converge (or "go together") as measures of life satisfaction.

The case for the construct validity of life satisfaction can be made even more strongly when the measures are shown to have discriminant validity. As can be seen in Table 5.1, the correlations between the Satisfaction with Life Scale (SWLS) and Positive Affect (.42) and between the 5-item Life Satisfaction measure (LS-5) and Positive Affect (.47) are lower. These findings show that life satisfaction measures do not correlate as well with a measure of another theoretical construct—namely, positive affect. The lower correlations between the life satisfaction tests and the positive affect test indicate that *different* constructs are being measured. Thus, there is evidence for *discriminant validity* of the life

BOX 5.2

COLLEGE STUDENTS' VALUES REVISITED: RELIABILITY AND VALIDITY

When describing the successive independent samples design, we presented data that suggest that first-year college students' values are oriented toward "being well-off financially" rather than "developing a meaningful philosophy of life." Now we can ask, "Do these two questions assess students' desire for meaning and purpose in their life in a reliable and valid manner?"

Reliable and valid measurement of a psychological construct such as "meaning and purpose in life" requires more than two questions and, in fact, data from the 2006 sample of students suggest that students are not concerned simply with financial goals (Bryant & Astin, 2006). Here are the percentages for other items endorsed by students as "essential" or "very important":

Attaining wisdom	77%
Becoming a more loving person	67%
Seeking beauty in my life	54%
Improving the human condition	54%
Attaining inner harmony	49%
Finding answers to mysteries of life	45%
Developing a meaningful philosophy of life	42%

Results for these additional items show that students clearly are interested in developing a meaningful life in ways other than pursuing purely financial goals. The item "developing a meaningful philosophy of life" seems to show weaker agreement or convergent validity with the other items, perhaps making it a poor item to represent the broader construct of meaning and purpose in life.

Could there be a problem with the wording "meaningful philosophy of life"? Students may have been less clear about the meaning of this item than the more concrete life goals indicated by the other items. Reliable and valid measurement requires clear, unambiguous questions—a topic addressed in the next section.

satisfaction measures because they seem to "discriminate" life satisfaction from positive affect—being satisfied with one's life is not the same as general happiness. The construct validity of life satisfaction gains support in our example because there is evidence for both convergent validity and discriminant validity. Box 5.2 provides another example of reliable and valid measurement.

Constructing a Questionnaire

- Constructing a questionnaire involves deciding what information should be sought and how to administer the questionnaire, writing a draft of the questionnaire, pretesting the questionnaire, and concluding with specifying the procedures for its use.
- The wording of questionnaires should be clear and specific using simple, direct, and familiar vocabulary.
- The order in which questions are asked on a questionnaire needs to be considered seriously because the order can affect respondents' answers.

Steps in Preparing a Questionnaire

Constructing a questionnaire that will yield reliable and valid measurements is a challenging task. In this section we suggest a series of steps that can help you meet this challenge, especially if you are constructing a questionnaire for the first time as part of a research project.

1 Decide what information should be sought.
2 Decide how to administer the questionnaire.

3 Write a first draft of the questionnaire.
4 Reexamine and revise the questionnaire.
5 Pretest the questionnaire.
6 Edit the questionnaire and specify the procedures for its use.

Step 1. The warning "Watch out for that first step!" is appropriate here. The first step in questionnaire construction—deciding what information is to be sought—should actually be the first step in planning the survey as a whole. This decision, of course, determines the nature of the questions to be included in the questionnaire. It is important to predict the likely results of a proposed questionnaire and decide whether these "findings" would answer the questions of the study. Surveys are frequently done under considerable time pressure, and inexperienced researchers are especially prone to impatience. A poorly conceived questionnaire, however, takes as much time and effort to administer and analyze as does a well-conceived questionnaire. The difference is that a well-constructed questionnaire leads to interpretable results. The best that can be said for a poorly designed one is that it is a good way to learn the importance of careful deliberation in the planning stages.

Step 2. The next step is to decide how to administer the questionnaire. For example, will it be self-administered, or will trained interviewers be using it? This decision is determined primarily by the survey method that has been selected. For instance, if a telephone survey is to be done, trained interviewers will be needed. In designing the questionnaire, one should also consider using items that have been prepared by other researchers. For example, there is no reason to develop your own instrument to assess racial prejudice if a reliable and valid one is already available. Besides, if you use items from a questionnaire that has already been used, you can compare your results directly with those of earlier studies.

Step 3. If you decide that no available instrument suits your needs, you will have to take the third step and write a first draft of your own questionnaire. Guidelines concerning the wording and ordering of questions are presented later in this section.

Step 4. The fourth step in questionnaire construction—reexamining and rewriting—is an essential one. Questions that appear objective and unambiguous to you may strike others as slanted or ambiguous. It is most helpful to have your questionnaire reviewed by experts, both those who have knowledge of survey research methods and those with expertise in the area on which your study is focused. For example, if you are doing a survey of students' attitudes toward the campus food service, it would be advisable to have your questionnaire reviewed by the campus food-service director. When you are dealing with a controversial topic, it is especially important to have representatives of both sides of the issue screen your questions for possible bias.

Step 5. By far the most critical step in the development of an effective questionnaire is to do a pretest. A pretest involves actually administering the questionnaire to a small sample of respondents under conditions

similar to those anticipated in the final administration of the survey. Pretest respondents must also be typical of those to be included in the final sample; it makes little sense to pretest a survey of nursing home residents by administering the questionnaire to college students. There is one way, however, in which a pretest does differ from the final administration of the survey. Respondents should be interviewed at length regarding their reactions to individual questions and to the questionnaire as a whole. This provides information about potentially ambiguous or offensive items.

The pretest should also serve as a "dress rehearsal" for interviewers, who should be closely supervised during this stage to ensure that they understand and adhere to the proper procedures for administering the questionnaire. If major changes have to be made as a result of problems discovered during the pretest, a second pretest may be needed to determine whether these changes solved the problems.

Step 6. After pretesting is completed, the final step is to edit the questionnaire and to specify the procedures to be followed in its final administration. To reach this final step successfully, it is important to consider guidelines for the effective wording of questions and for the ordering of questions.

Guidelines for the Effective Wording of Questions Lawyers have long known that how a question is phrased has great impact on how that question is answered. Survey researchers need to be equally conscious of this principle. This point is illustrated in a study that examined people's opinions about allocating scarce vaccines during a hypothetical flu epidemic (Li, Vietri, Galvani, & Chapman, 2010). These researchers found that respondents' decisions about vaccine allocation (in effect, who would live and who would die) were affected by whether vaccination policies were written in terms of "saving lives" *versus* "lives lost." Thus, the way the questions were worded influenced how respondents judged the value of people's lives. In a typical survey, only one wording is used for each question so, unfortunately, the influence of the wording of questions in a given survey can almost never be determined precisely.

Clark and Schober (1992) point out that respondents presume that the meaning of a question is obvious. This has important implications. For instance, when a question includes a vague word, respondents may interpret the word in various ways according to their individual biases and their own ideas of what is "obvious." Thus, words like "few" or "usually," or terms such as "global warming," may be interpreted differently by different individuals. Respondents also tend to assume that words in a survey are used in the same way as in their subculture or culture. A recent example in popular culture is figuring out whether "bad" means "good." Clark and Schober (1992) cite as an example a surveyor who wanted to ask Mexican residents in the Yucatán the question "How many children do you have?" When translated into Spanish, the surveyor used the word *niños* for children, but villagers in this area of Mexico treated *niños* as including living children and children who have died. Respondents also may reasonably assume that if the surveyor asks a question, then it is one that the respondent can answer. This assumption can lead respondents to give answers

to questions that have no (valid) answers! For example, when asked to give opinions about nationalities that didn't actually exist, respondents nevertheless gave opinions.

Although it's clear that question wording in surveys can pose problems, the solution is less clear. *At a minimum, the exact wording of critical questions should always be reported along with the data describing respondents' answers.* The problem of the potential influence of the wording of questions is yet another illustration of why a multimethod approach is so essential in investigating behavior.

Survey researchers usually choose from two general types of questions when writing a questionnaire. The first type is a *free-response* (open-ended) question and the second type is a *closed* (multiple-choice) question. Free-response questions, like the essay questions on a classroom test, merely specify the area to be addressed in a response. For example, the question "What are your views on legal abortion?" is a free-response question. By contrast, closed questions provide specific response alternatives. "Is police protection very good, fairly good, neither good nor bad, not very good, or not good at all?" is a closed question about the quality of police protection in a community.

The primary advantage of free-response questions is that they offer the respondent greater flexibility than closed questions. However, this advantage is often more than offset by the difficulties that arise in recording and scoring responses to free-response questions. For example, extensive coding is frequently necessary to summarize rambling responses to free-response questions. Closed questions, on the other hand, can be answered more easily and quickly and fewer scoring problems arise. It is also much easier to summarize responses to closed questions because the answers are readily comparable across respondents. A major disadvantage of closed questions is that they reduce expressiveness and spontaneity. Further, respondents may have to choose a less-than-preferred response because no presented alternative really captures their views. Hence, the responses obtained may not accurately reflect the respondents' opinion.

Regardless of the type of question used, the *vocabulary should be simple, direct, and familiar to all respondents.* Questions *should be as clear and specific as possible. Double-barreled questions should be avoided.* An example of a double-barreled question is "Have you suffered from headaches and nausea recently?" A person may respond "no" if both symptoms have not occurred at exactly the same time or may respond "yes" if either symptom has occurred. The solution to the problem of double-barreled questions is a simple one—rewrite them as separate questions.

Survey questions should be as short as possible without sacrificing the clarity of the questions' meaning. Twenty or fewer words should suffice for most survey questions. *Each question should be carefully edited for readability and should be phrased in such a way that all conditional information precedes the key idea.* For example, it would be better to ask, "If you were forced to leave your present job, what type of work would you seek?" than to ask, "What type of work would you seek if you were forced to leave your present job?"

Leading or loaded questions should also be avoided in a questionnaire. *Leading* questions take the form "Most people favor the use of nuclear energy. What

do you think?" To avoid bias, it is better to mention all possible perspectives or to mention none. A survey question about attitudes toward nuclear energy could read, "Some people favor the use of nuclear energy, some people oppose the use of nuclear energy, and some people have no opinion one way or the other. What do you think?" or "What do you think about the use of nuclear energy?" *Loaded* questions are questions that contain emotion-laden words. For example, terms such as "radical" and "racist" should be avoided. To guard against loaded questions, it is best to have your questionnaire reviewed by individuals representing a range of social and political perspectives.

Finally, when using multiple items to assess a construct, it's important to word some of the items in the opposite direction to avoid problems associated with *response bias.* The potential for response bias exists when respondents use only extreme points on rating scales, or only the midpoint, or when respondents agree (or disagree) with every item. For example, an assessment of emotional well-being might include the following items:

My mood is generally positive.

1	2	3	4	5
Strongly disagree				Strongly agree

I am often sad.

1	2	3	4	5
Strongly disagree				Strongly agree

Respondents with a response bias in which they always agree with statements might circle "5" on both scales, resulting in an unreliable assessment of emotional well-being. More consistent responding would require participants to use the opposite end of the self-rating scale. Responses to these reversed items are "reverse-scored" (1 = 5, 2 = 4, 4 = 2, 5 = 1) when participants' responses are summed to derive a total score for emotional well-being.

In summary, good questionnaire items should

—use vocabulary that is simple, direct, and familiar to all respondents;
—be clear and specific;
—not involve leading, loaded, or double-barreled questions;
—be as short as possible (20 or fewer words);
—present all conditional information prior to the key idea;
—avoid potential response bias; and
—be checked for readability.

Ordering of Questions The order of the questions in a survey requires careful consideration. The first few questions set the tone for the rest of the questionnaire, and determine how willingly and conscientiously respondents will work on subsequent questions. For self-administered questionnaires, it is best to begin with the most interesting set of questions in order to capture the respondents' attention. Demographic data should be obtained at the end of a

self-administered questionnaire. For personal or telephone interviews, on the other hand, demographic questions are frequently asked at the beginning because they are easy for the respondent to answer and thus bolster the respondent's confidence. They also allow time for the interviewer to establish rapport before asking questions about more sensitive matters.

The order in which particular questions are asked can have dramatic effects, as illustrated in a study by Schuman, Presser, and Ludwig (1981). They found differential responding depending on the order of two questions concerning abortion, one general and one specific. The general question was "Do you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children?" The more specific question was "Do you think it should be possible for a pregnant woman to obtain a legal abortion if there is a strong chance of a serious defect in the baby?" When the general question was asked first, 60.7% of respondents said "yes," but when the general question followed the specific question, only 48.1% of respondents said "yes." The corresponding values for the specific question were 84% and 83% agreement in the first and second positions, respectively. The generally accepted method for dealing with this problem is to use *funnel questions,* which means starting with the most general question and moving to specific questions pertaining to a given topic.

The final aspect of the ordering of survey questions that we will consider is the use of *filter questions*—general questions asked of respondents to find out whether they need to be asked specific questions. For example, the question "Do you own a car?" might precede a series of questions about the costs of maintaining a car. In this instance, the respondents would answer the specific questions only if their response to the general question was "yes." If that answer was "no," the interviewer would not ask the specific questions (in a self-administered questionnaire, the respondent would be instructed to skip that section). When the filter questions involve objective information (e.g., "Are you over 65?"), their use is relatively straightforward. Caution must be exercised, however, in using behavioral or attitudinal questions as filter questions. Smith (1981) first asked respondents whether they approved of hitting another person in "any situations you can imagine." Logically, a negative response to this most general question should imply a negative response to any specific questions. Nonetheless, over 80% of the people who responded "no" to the general question then reported that they approved of hitting another person in specific situations, such as in self-defense. Although findings such as this suggest that filter questions should be used cautiously, the need to demand as little of the respondents' time as possible makes filter questions an essential tool in the design of effective questionnaires.

A well-conducted survey is an efficient way to accomplish the research goals of description and prediction. When distributed to dozens if not hundreds of individuals, even a modest-sized questionnaire can quickly generate many thousands of responses to individual items. And, as we have seen, by using the Internet, researchers can literally obtain millions of responses in a short

period of time. But there is a catch! How does one deal with this multitude of responses? The answer is: By careful planning!

Data analysis of responses obtained from questionnaires must be considered prior to writing the survey items. Will open-ended questions be used? Is the goal mainly descriptive; for example, are proportions or percentages of events in a population of primary interest? Is the goal correlational, for example, relating responses on one question to those of another? Will respondents use a yes–no response format? A yes–maybe–no format? Self-report scales? These response formats provide different kinds of data. As you have learned, qualitative data in the form of open-ended responses will require rules for coding and methods for getting intercoder reliabilities. Categorical data obtained from a yes–no format yield nominal data, whereas scales are typically assumed to provide interval data (see Chapter 4 for comments on types of scales). These types of data require different approaches for statistical analysis.

It is important to anticipate the likely results of the proposed questionnaire and then to decide whether these "findings" will answer the research questions. When "predicting" your results, you will want to make sure that the results can be analyzed appropriately. In other words, *you should have an analysis plan prior to conducting the survey*. During the planning stage, we suggest that you consult with experienced survey researchers regarding the correct statistical analyses.

Once again we refer you to Chapters 11 and 12 of this textbook to gain (or regain) familiarity with statistical procedures. Should your interest in conducting a survey lead you to look for relationships (correlations) among categorical (nominal) variables, you will need to go beyond this textbook. *The appropriate statistical analysis for examining relationships between nominal variables is the chi-square test of contingency.* An introduction to this test is found in nearly all introductory statistics books (e.g., Zechmeister & Posavac, 2003). If you are going to correlate responses to interval scales, then a Pearson Product-Moment correlation (r) is appropriate. This type of analysis was introduced in Chapter 4 when we discussed interobserver reliability. We will have more to say about correlational analyses toward the end of this chapter. Procedures for calculating a Pearson r are found in Chapter 11.

Thinking Critically About Survey Research

Correspondence Between Reported and Actual Behavior

- Survey research involves reactive measurement because individuals are aware that their responses are being recorded.
- Social desirability refers to pressure that respondents sometimes feel to respond as they "should" believe rather than how they actually believe.
- Researchers can assess the accuracy of survey responses by comparing these results with archival data or behavioral observations.

Regardless of how carefully survey data are collected and analyzed, the value of these data depends on the truthfulness of respondents' answers to the survey questions. Should we believe that people's responses on surveys reflect their true thoughts, opinions, feelings, and behavior? The question of the truthfulness of verbal reports has been debated extensively, and no clear-cut conclusion has emerged. In everyday life, however, we regularly accept the verbal reports of others as valid. If a friend tells you she enjoyed reading a certain novel, you may ask why, but you do not usually question whether the statement accurately reflects your friend's feelings. There are some situations in everyday life, however, when we *do* have reason to suspect the truthfulness of someone's statements. When looking for a used car, for instance, we might not always want to trust the "sales pitch" we receive. Generally, however, we accept people's remarks at their face value unless we have reason to suspect otherwise. We apply the same standards to the information we obtain from survey respondents.

By its very nature, survey research involves reactive measurement. Respondents know their responses are being recorded, and they may also suspect their responses may prompt some social, political, or commercial action. Hence, pressures are strong for people to respond as they "should" believe, not as they actually believe. The term often used to describe these pressures is **social desirability** (the term "politically correct" refers to similar pressures). For example, if respondents are asked whether they favor giving help to the needy, they may say "yes" because they believe this is the most socially acceptable attitude to have. In survey research, as was true with observational research, the best protection against reactive measurement is to be aware of its existence.

Key Concept

Sometimes researchers can examine the accuracy of verbal reports directly. For example, Judd, Smith, and Kidder (1991) describe research by Parry and Crossley (1950) wherein responses obtained by experienced interviewers were subsequently compared with archival records for the same respondents kept by various agencies. Their comparisons revealed that 40% of respondents gave inaccurate reports to a question concerning contributions to United Fund (a charitable organization), 25% reported they had registered and voted in a recent election (but they did not), and 17% misrepresented their age. A pessimist might find these figures disturbingly high, but an optimist would note that a majority of respondents' reports were accurate even when social desirability pressures were high, as in the question pertaining to charitable contributions.

Another way researchers can assess the accuracy of verbal reports is by directly observing respondents' behavior. An experiment done by Latané and Darley (1970) illustrates this approach. They found that bystanders are more likely to help a victim when the bystander is alone than when other witnesses are present. Subsequently, a second group of participants was asked whether the presence of others would influence the likelihood they would help a victim. They uniformly said that it would not. Thus, individuals' verbal reports *may not* correspond well to behavior (see Figure 5.6). Research findings such as these should make us extremely cautious of reaching conclusions about people's behavior solely on the basis of verbal reports. Of course, we should

FIGURE 5.6 How people say they would respond to this type of situation does not always match what they actually do.

be equally cautious of reaching conclusions about what people think solely on the basis of direct observation of their behavior. The potential discrepancy between observed behavior and verbal reports illustrates again the wisdom of a multimethod approach in helping us identify and address potential problems in understanding behavior and mental processes.

Correlation and Causality

- When two variables are related (correlated), we can make predictions for the variables; however, we cannot, simply knowing a correlation, determine the cause of the relationship.
- When a relationship between two variables can be explained by a third variable, the relationship is said to be "spurious."
- Correlational evidence, in combination with a multimethod approach, can help researchers identify potential causes of behavior.

Surveys are often used in correlational research, and correlational research is an excellent method for meeting the scientific goals of description and prediction. For example, studies demonstrating correlations between physical health and psychological well-being allow researchers to make predictions regarding health-related problems.

Correlational evidence allows researchers to make predictions for the correlated variables. However, the familiar maxim, "Correlation does not imply causation," reminds us that our ability to make causal inferences based solely on a correlation between two variables is very limited. For instance, there is a reliable correlation between being outgoing (socially active) and being satisfied with one's life (Myers & Diener, 1995). Based on this correlation alone, however, we could not argue convincingly that being more outgoing and socially active *causes* people to be more satisfied with their lives. Although it is possible that being outgoing causes people to be more satisfied, the "reverse" causal relationship also may be true: Being satisfied with life may cause people to be more outgoing and socially active. The causal relationship could go either

way—being more outgoing causes greater life satisfaction or being more satisfied with life causes people to be more outgoing. It is impossible to determine the correct causal direction simply by knowing the correlation between the two variables.

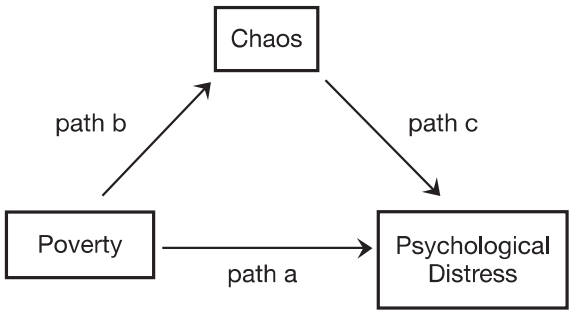
Not being able to determine the direction of the relationship in a correlation is only one challenge we face. It's possible there is another causal interpretation for the correlation between the two variables. For example, a third variable, number of friends, could cause people to be more outgoing *and* more satisfied with their lives. A correlation that can be explained by a third variable is called a **spurious relationship** (Kenny, 1979). In this particular example, "number of friends" is a possible third variable that could account for the relationship between being outgoing and being satisfied with one's life. Individuals with more friends may be more likely to be outgoing and satisfied with life than people with fewer friends. This isn't to say that the original positive correlation between being outgoing and life satisfaction doesn't exist (it certainly does); it just means that other variables that were not measured (e.g., number of friends) may explain *why* the relationship exists.

Key Concept

It is extremely important to understand why it is not possible to make a causal inference based only on a correlation between two variables. It is equally important to recognize that correlational evidence can be very useful in identifying *potential* causes of behavior. Sophisticated statistical techniques can be used to help with causal interpretations of correlational studies. *Path analysis* is one sophisticated statistical technique that can be used with correlational data (Baron & Kenny, 1986; Holmbeck, 1997). Path analysis involves the identification of mediator variables and moderator variables. A *mediator* variable is a variable that is used to explain the correlation between two variables. A *moderator* variable is a variable that affects the direction or strength of the correlation between two variables.

Figure 5.7 illustrates an example of a mediating variable in a study of the effects of poverty on children's psychological adjustment (Evans, Gonnella, Marcynyszyn, Gentile, & Salpekar, 2005). Consistent with previous research, these investigators observed a correlation between their measures of poverty and psychological distress: the greater the poverty, the greater the distress among children (path *a* in Figure 5.7). Evans and his colleagues also proposed a mediating variable, *chaos,* to account for this relationship. They theorized that

FIGURE 5.7 An example of a mediating variable.



chaotic living conditions characterized by unpredictability, confusion, lack of structure, noise, overcrowding, and poor-quality housing can explain the relationship between poverty and children's psychological distress. This is shown in paths *b* and *c* in Figure 5.7.

Consistent with their predictions, the results of their study indicated that greater poverty was associated with greater chaos in the home (path *b*). Also, greater chaos was associated with greater psychological distress (path *c*). The final step in path analysis is to show that when the correlations between paths *b* and *c* are taken into account using a statistical procedure, the correlation observed initially for path *a* (between poverty and distress) becomes zero (i.e., no relationship). This is exactly what Evans and his colleagues found. Their path analysis allowed them to say that the relationship between poverty and children's distress can be explained by, or *is mediated by,* the degree of chaos in the home.

Although Evans and his colleagues did not describe potential moderating variables, we can offer a hypothetical illustration. Suppose the pattern of correlations observed in Figure 5.7 is different for boys compared to girls. We could hypothesize, for example, that the mediating effect of chaos exists only for boys and not for girls. In this case we would be arguing that the sex of the child, boy or girl, is a moderating variable—that is, it affects the direction or strength of the correlations among poverty, chaos, and psychological distress. Other potential moderating variables might include population density (e.g., urban *vs.* rural) and the extent of resilience in the children's personality (e.g., high *vs.* low resilient). Can you develop hypotheses for how the relationships among poverty, chaos, and psychological distress may differ based on these moderating variables?

Although correlational research is not an absolutely firm basis for making causal inferences, patterns of correlations observed in path analysis provide important clues for identifying causal relationships among variables. The next step for researchers who wish to make causal inferences is to conduct experiments, as described in Chapters 6–8. For example, a laboratory manipulation of chaos (e.g., unpredictable outcomes, noise) might cause different levels of distress among individuals from different economic backgrounds. This multimethod approach would help to provide converging evidence regarding the causal role of chaos in understanding the relationship between poverty and psychological adjustment.

Summary

Survey research provides an accurate and efficient means for describing people's characteristics (e.g., demographic variables) and their thoughts, opinions, and feelings. In addition, predictive relationships can be identified by assessing the covariation (correlation) among naturally occurring variables. Surveys differ in purpose and scope, but they generally involve sampling. Results obtained for a carefully selected sample are used to describe the entire population of interest. Surveys also involve the use of a predetermined set of questions, generally in the form of a questionnaire.

Sampling is a procedure whereby a specified number of elements are drawn from a sampling frame that represents an actual list of the possible elements in the population. Our ability to generalize from the sample to the population depends critically on the representativeness of the sample, the extent to which the sample has the same characteristics as the population. Representativeness is best achieved by using probability sampling rather than nonprobability sampling. In simple random sampling, the most common type of probability sampling, every element is equally likely to be included in the sample. Stratified random sampling is used when analysis of subsamples is of interest.

There are four general survey methods: mail surveys, personal interviews, telephone interviews, and Internet surveys. Mail surveys avoid problems of interviewer bias and are especially well suited for examining personal or embarrassing topics. Potential problems due to response rate bias are a serious limitation of mail surveys. Personal interviews and phone surveys usually have higher response rates and provide greater flexibility. The phone survey is frequently used for brief surveys. Internet surveys are efficient and cost effective and open new opportunities for survey researchers; however, they are also prone to sample biases and raise both methodological and ethical issues primarily due to the lack of control over the research environment.

Survey research is carried out according to an overall plan called a research design. There are three survey-research designs: the cross-sectional design, the successive independent samples design, and the longitudinal design. Cross-sectional designs focus on describing the characteristics of a population or the differences between two or more populations at one point in time. Describing changes in attitudes or opinions over time requires the use of successive independent samples or longitudinal designs. The longitudinal design is generally preferred because it allows the researcher to assess changes for specific individuals and avoids the problem of noncomparable successive samples.

The primary instrument for survey research is the questionnaire. Questionnaires can be used to measure demographic variables and to assess people's preferences or attitudes. In order to construct questionnaires that will yield reliable and valid measurements, researchers must decide what information should be sought and how to administer the questionnaire, and what order of questions will be most effective. Most importantly, questions must be written so that they are clear, specific, and as unambiguous as possible.

Survey results, like those of other verbal reports, can be accepted at face value unless there is reason to do otherwise, such as pressures on respondents to give socially desirable responses. People's behavior does not always conform to what they say they would do, so survey research will never replace direct observation. However, survey research does provide an excellent way to begin to examine people's attitudes and opinions.

The greatest challenge in interpreting correlational evidence is understanding the relationship between correlation and causality. A correlation between two variables is not sufficient evidence to demonstrate a causal relationship between the two variables. Correlational evidence can contribute, however, to

identifying causal relationships when used in combination with sophisticated statistical techniques (such as analyses of mediators and moderators in path analysis) and the multimethod approach.

Key Concepts

correlational research 138
population 141
sample 142
representativeness 143
selection bias 143
nonprobability sampling 144
probability sampling 144
simple random sampling 145
stratified random sampling 147
response rate bias 149
interviewer bias 150
cross-sectional design 154
successive independent samples design 155
longitudinal design 158
questionnaire 162
social desirability 174
spurious relationship 176

Review Questions

1. Briefly identify the goal of survey research and how correlations are used within survey research.
2. Describe the information you would examine to determine whether survey results are biased because the sponsoring agency of the survey has a vested interest in how the results turn out.
3. What two characteristics do surveys have in common regardless of the purpose for which the survey has been done?
4. Explain why there is likely to be a serious threat to the interpretability of the results of a survey when a convenience sample is used.
5. Explain the relationship between the homogeneity of the population from which a sample is to be drawn and the size of a sample needed to ensure representativeness.
6. Explain why you would choose to use a mail survey, personal interviews, telephone interviews, or an Internet survey for your survey-research project.
7. Explain why it is not possible to conclude a sample does not represent a population simply by knowing that the response rate was 50%.
8. What are the major advantages and disadvantages of Internet surveys?
9. Describe the relationship that would need to exist among the samples in a successive independent samples design in order to be able to interpret population changes in attitudes over time.
10. You are interested in assessing the direction and extent of change over time in the opinions of individual respondents. Identify the survey-research design you would choose, and explain why you would make this choice.
11. Describe one method for determining the reliability and one method for determining the validity of a self-report measure.
12. Describe three factors that affect the reliability of self-report measures in survey research.
13. How would you respond if someone told you that survey results were useless because people do not respond truthfully to questions on surveys?
14. Explain why "correlation does not imply causation," and explain how correlational evidence can be useful in identifying potential causes of behavior.
15. Define *mediator* and *moderator* and provide an example of each.

CHALLENGE QUESTIONS

1 Survey research is difficult to do well, and this can be especially the case when the topic is people's sexual attitudes and practices. For a book focusing in part on women's sexuality, an author mailed 100,000 questionnaires to women who belonged to a variety of women's groups in 43 states. These groups ranged from feminist organizations to church groups to garden clubs. The author's questionnaire included 127 essay questions. The author received responses from 4,500 women.

Findings in this survey included that 70% of respondents married 5 years or more reported having extramarital affairs and that 95% of respondents felt emotionally harassed by the men they love.

A The final sample in this study is large (4,500). Is this sufficient to ensure the representativeness of the sample? If not, what potential survey-research problem could lessen the sample's representativeness?

B Is it possible on the basis of your response to Part A of this question to argue that any conclusions drawn by the author from her data are incorrect? What could you do to determine whether the results are correct?

2 Two different national organizations that conduct research on higher education did independent surveys asking faculty how well prepared they thought their students were. The results of these two surveys drew attention when they were reported in the *Chronicle of Higher Education* because the findings from the two surveys were very different. Researchers from Research Foundation A found that nearly 75% of professors said that their students were "seriously underprepared." Researchers from Research Foundation B found that only 18.8% of the faculty they surveyed said that their students were "not at all prepared." Survey-research findings can be expected to vary from one survey to another, but the large discrepancy found in these two surveys could make one wonder about the reliability and credibility of survey findings. Before reaching this conclusion, it is useful to consider several details of the two surveys. [*Note:* This question is based on a report from the *NCRIPTAL Update,* Spring 1990, Vol. 3, No. 1, pp. 2–3.]

A *Who was asked?* The original sample for Foundation A included 10,000 college professors who taught undergraduate and graduate students in all types of institutions. Of the original sample, 54.5% responded. Foundation B omitted research universities (25% of Foundation A's sample). Foundation B had a final sample of 2,311 (62% response rate). Approximately 90% of the final sample were teaching introductory-level students. *How might the characteristics of the samples surveyed by Foundations A and B affect the findings obtained in the two surveys?*

B *What was asked?* Foundation A asked its respondents: "The undergraduates with whom I have close contact are seriously underprepared in basic skills such as those required for written and oral communication." The responses for this statement were: strongly agree, agree with reservations, neutral, disagree with reservations, and disagree. Foundation B asked its respondents: "In their background preparation, students who enroll in this course are most typically. . . ." The response choices were: not at all prepared, somewhat prepared, very well prepared, and extremely well prepared. *How might the nature of these questions affect the findings obtained in the two surveys?*

C How were the results reported? The findings for the Foundation A survey (75% of students seriously underprepared) were reported in the Chronicle by combining the response categories "strongly agree" and "agree with reservations." The findings for the Foundation B survey (18.8% of students not at all prepared) represented only respondents who chose the "not at all prepared" response category. *How do you think the results might look if the Foundation A estimate included only the respondents who chose the "strongly agree" response?*

3 A task force has been established at a small liberal arts college under the direction of the dean of students to examine the quality of students' experiences on their campus. The task force decided to do a survey to determine students' knowledge of and their perceptions of the fairness of the judicial system used to enforce the rules in the living units on campus. The questionnaire for the survey included personal questions asking students to describe their own experiences when they had violated college policies or when they had known other students who had violated college policies. A stratified random sample was drawn from the registrar's list of full-time students living on and off campus. The sample size was 400 on a campus with 2,000 full-time students. Questionnaires were returned by 160 students for a response rate of 40%. One important finding from the survey was that over a third of the respondents rated the judicial system as unfair. The task force

met to decide whether to include the survey findings such as this one in its final report to the dean of students.

A Was the initial sample of 400 students likely to be representative of the population of 2,000 full-time students? Why or why not?

B Identify a potential survey research problem that could be present in this study that would lead the task force to be concerned that the final sample was not representative of the population of 2,000 students.

C Using only the evidence that the response rate for the survey was 40%, the task force concluded that the final sample was not representative of the population of students. They further decided that the ratings of the judicial system as unfair by more than a third of the students was an incorrect overestimate. Do you agree that the finding represents an incorrect estimate? Why or why not?

D While the task force was meeting to discuss their final report, one member of the task force expressed the opinion that students' responses were unlikely to have been truthful and so the results of the survey were useless and should not be reported at all. The director of the task force calls on you to respond to this statement. What would you say?

4 As an intern with the alumni relations office at a small college, one of your assignments is to help develop a survey-research project. The college is interested in finding out about the alumni's attitudes toward their academic and extracurricular experiences while enrolled in college. The director also wants to include questions to assess the alumni's opinions about the different activities the college sponsors for them (e.g., reunions) and how they prefer to be kept informed about issues and activities on campus (e.g., newsletters, e-mails, postings on the college website). One of the major goals of the survey-research project is to determine how the attitudes of alumni change 1, 5, or 10 years after graduation.

A The first step is to select the survey-research design for the project. Describe the two designs that can be used to measure changes in attitudes over time. Outline how each of these designs would be implemented for this project, and identify the advantages and possible limitations of each design.

B The second step is to select the survey-research method for the project. Members of the planning committee proposed three different approaches: (1) select a random sample of alumni from the alumni relations office list and use a phone survey to administer the questionnaire; (2) send an e-mail to a random sample of alumni that includes a link to an Internet site where alumni can complete the questionnaire; (3) post an announcement about the survey and a link to the questionnaire on the college's website with the request that all alumni visiting the website complete the questionnaire. Describe to the committee the advantages and limitations of each approach, and provide a recommendation and rationale for which approach you think would be best.

C The third step is to prepare the questionnaire. Describe the different formats that can be used to write the questionnaire items and prepare an example of a free-response (open-ended) and closed (multiple-choice) question. Use these examples to describe the advantages and disadvantages of each type of question.

Answer to Stretching Exercise I

1 D **2** B **3** A **4** C

Answer to Stretching Exercise II

1 The first student researcher is proposing a stratified random sample in which 100 "Greek" and 100 "independent" students are sampled. In this plan the equal-sized strata would have representative samples for each stratum. A potentially serious flaw of this plan is that the overall sample would not represent the proportions of Greeks and independents in the population (25% and 75%, respectively). This would result in a biased sample because Greeks would be systematically overrepresented in the survey. The second student researcher is proposing a simple random sample of 100 students from the campus population. While this is likely to lead to a more representative sample, it will probably result in too few respondents in the "Greek" category (we'd expect about 25 Greeks) to adequately represent their viewpoint.

2 A preferred sampling plan would use a stratified random sample in which the sample sizes for Greeks and independents are proportional to the population values. With 200 students in the sample, you would select 150 students from the sampling frame of independent students and 50 students from the sampling frame of Greek students.

Answer to Challenge Question 1

A In general, larger sample sizes do make it more likely that the sample will be representative. The problem in this study is that the final sample (though large) represents a low response rate from the original sample of 100,000 (4.5%). The low response rate and the topic of the survey make it likely that only women who were very motivated to complete the survey responded. It is unlikely the sample of 4,500 women represents the entire population of women.

B The low response rate does not make it possible to argue that the conclusions drawn by the author are incorrect. Neither can the author argue on the basis of this sample that the conclusions are correct. We simply cannot know based on this evidence whether the conclusions are correct or incorrect. There is at least one good way to determine if the results of this survey are correct. You would need to obtain from the literature the results of one or more surveys on women's sexual attitudes and practices. It would be essential that these other surveys had used representative samples of women. Then you would compare the results of this survey with those of the other surveys. Only if the results of the present survey corresponded to those of the surveys with the representative samples would we consider the results of the present survey correct. Of course, you could also carry out your own survey, one that avoids the problems that are present in this survey, and determine whether your results are similar to those of this author-researcher!