Nonexperimental Methods: Descriptive Methods, Correlational Studies, Ex Post Facto Studies, Surveys and Questionnaires, Sampling, and Basic Research Strategies



Descriptive Methods

- · Archival and Previously Recorded Sources of Data
- Naturalistic Observation
- Choosing Behaviors and Recording Techniques

Correlational Studies

- The Nature of Correlations
- Correlational Research

Ex Post Facto Studies

Surveys, Questionnaires, Tests, and Inventories

- Surveys and Questionnaires
- Tests and Inventories

Sampling Considerations and Basic Research Strategies

Sampling • Basic Research Strategies

Unlike the qualitative research techniques that we covered in Chapter 3, which also are nonexperimental procedures, the nonexperimental methods covered in this chapter will have a more narrowly focused research problem and will use statistical procedures to analyze the data.

Descriptive Methods

Because several of these methods do not involve the manipulation of any variables, they are called **descriptive research methods**. When we use descriptive methods, we can only speculate about causation that may be involved.

Archival and Previously Recorded Sources of Data

Descriptive research methods Research methods that do not involve the manipulation of an independent variable.

In some instances researchers may not gather their own data; they may answer their research question by using data recorded by other individuals for other purposes. For example, public health and census data may be analyzed years later to answer questions about socioeconomic status, religion, or political party affiliation. In some instances the records and data you need to consult are stored in a central location. The Archives of the History of American Psychology were established at the University of Akron in 1974 for conducting research on the history of psychology. Although the letters, documents, and photographs contained in this collection were originally produced for numerous reasons, they are now used by researchers interested in answering questions about the history of our discipline.

You also can access several archival sources online. For example, the General Social Survey (GSS) has been conducted almost annually since 1972 by the National Opinion Research Center. The responses of more than 35,000 respondents to a wide range of socially relevant questions are available at http://www.icpsr.umich.edu/gss/home.htm. The number of research questions that can be addressed with this huge body of information seems endless. There is no charge for accessing this site; we encourage you to peruse it.

Not all sources of previously recorded data are stored in an archive or library or are available online for our use. There are numerous real-life sources. For example, Laura Marzola, a student at King's College (Wilkes-Barre, PA), and her faculty supervisor, Charles Brooks, examined back issues of the *Psi Chi Journal of Undergraduate Research* to determine which journals student researchers cited most frequently in their published articles (Marzola & Brooks, 2004). Similarly, Jennifer Salhany, a student at St. John's University (Staten Island, NY), and her faculty supervisor, Miguel Roig, used college and university handbooks and catalogs found on the Internet to determine the prevalance and nature of academic dishonesty policies (Salhany & Roig, 2004).

Moreover, if you are interested in differences in sexual preoccupation between men and women, you might choose to examine graffiti on the walls of public restrooms.



The use of archival and previously recorded data is certainly in line with being a psychological detective. You are putting together bits and pieces of data to answer research questions. Unfortunately, several problems are associated with this approach to gathering information. Consider conducting this type of research and see what problems you can discover.

Potential Problems There are several potential problems associated with using archival and previously recorded sources of data. First, unless you are dealing with the papers and documents of a few clearly identified individuals, you will not know exactly who left the data you are investigating. Not knowing the participants who make up your sample will make it difficult to understand and generalize your results. Consider the graffiti example. You may choose to record graffiti from restrooms on your campus. Who created the graffiti? Was it created by a representative sample of students? Although common sense may tell you that the sample of graffiti writers is not representative of students on your campus, let alone college students in general, you do not know. Your ability to make statements other than those that merely describe your data is severely limited.

Second, the participants may have been selective in *what* they chose to write. Clearly, this consideration may be important in our graffiti example, particularly if we are evaluating the presence of sexual comments. What others chose to record may drastically influence our

conclusions in other instances. For example, until recently what we knew about Wilhelm Wundt, the founder of scientific psychology, was provided by E. B. Titchener, who originally translated Wundt's books and papers from German into English. Unfortunately, Titchener chose to misrepresent Wundt on several occasions; hence, our impression of Wundt may be severely distorted (Goodwin, 2005). Fortunately, Wundt's original writings are still available for retranslation and examination. Even if his works are retranslated, we will still face the possible problem that Wundt may have omitted things from his writings which he did not want to share with others. Whenever archival or previously recorded sources are used, you cannot avoid this problem of *selective deposit*.

A third problem with this type of data concerns the survival of such records. In our study of graffiti it will be important to know something about the cleaning schedule for the restroom(s) we are observing. Are they scrubbed clean each day, or are graffiti allowed to accumulate? In this example the data in which you are interested will probably not have a very high survival rate. Printed materials may not fare much better. During the 1920s John Watson and his call to behaviorism made psychology immensely popular in the United States; the number of journal articles and books written during this period attests to its popularity. It was only recently, however, that researchers discovered a very popular magazine, *Psychology: Health, Happiness, and Success,* which was published from 1923 to 1938 (Benjamin, 1992). Why does mystery surround this once-popular magazine? The problem had to do with the type of paper on which the magazine was printed. The high acid content of the paper led to rapid disintegration of these magazines; hence, only a precious few have survived.

Comparisons with the Experimental Method Certainly, the researcher can gain valuable information from archival and prerecorded sources. However, we must be aware of the problems of a nonrepresentative sample, data that are purposely not recorded, and data that have been lost. A comparison of using this technique with conducting an experiment reveals other weaknesses in addition to these limitations. Because we examined data and documents that were produced at another time under potentially unknown circumstances, we are not able to exercise any control with regard to the gathering of these data. We are therefore unable to make any type of cause-and-effect statement; the best we can do is speculate about what might have occurred.

These concerns notwithstanding, this type of research can yield interesting and valuable results. For example, much of what you read in a history of psychology text is the product of archival research. In the next section we examine methods in which we observe the phenomenon of interest firsthand.

Naturalistic Observation

As we saw in Chapter 3, **naturalistic observation** involves seeking answers to research questions by observing behavior in the real world and is the hallmark of qualitative research. However, researchers also use naturalistic observation to collect numerical data and answer more focused research questions. For example, each spring, animal psychologists interested in the behavior of migrating sandhill cranes conceal themselves in Naturalistic observation Seeking answers to research questions by observing behavior in the real world.

camouflage blinds to observe the roosting behavior of these birds on the Platte River in central Nebraska. For another example, a researcher who is interested in the behavior of preschool children might go to a day-care center to observe the children and record their behavior.

The possibilities for conducting studies using naturalistic observation are limited only by our insight into a potential area of research. Regardless of the situation we choose, we have two goals in using naturalistic observation. The first goal should be obvious from the name of the technique: to describe behavior as it occurs in the natural setting without the artificiality of the laboratory. If the goal of research is to understand behavior in the real world, what better place to gather research data than in a natural setting? The second goal of naturalistic observation is to describe the variables that are present and the relations among them. Returning to our sandhill crane example, naturalistic observation may provide clues concerning why the birds migrate at particular times of the year and what factors determine the length of stay in a certain area.

In a naturalistic observation study, it is important that the researcher not interfere with or intervene in the behavior being studied. For example, in our study of preschoolers, the observer should be as inconspicuous as possible. For this reason, the use of one-way mirrors, which allow researchers to observe without being observed, is popular.



Why should the researcher be concealed or unobtrusive in a study using naturalistic observation?

The main reason the researcher must be unobtrusive in studies using naturalistic observation is to avoid influencing or changing the behavior of the participants being observed. The presence of an observer is not part of the natural setting for sandhill cranes or

preschoolers; they may well behave differently in the presence of observers.

Reactance or **reactivity effect** The finding that participants respond differently when they know they are being observed. The **reactance** or **reactivity effect** refers to the biasing of the participants' responses because they know they are being observed. Perhaps the most famous example of a reactivity effect occurred in a study conducted at the Western Electric Company's Hawthorne plant located on the boundary between Chicago and Cicero, Illinois, in the late 1930s (Roethlisberger & Dickson, 1939). The purpose of the research was to determine the effects of factors such as working hours and lighting on

productivity. When researchers compared the productivity of the test participants to that of the general plant, an unusual finding emerged. The test participants produced at a higher rate, often under test conditions that were inferior to those normally experienced. For example, when the room lighting was reduced well below normal levels, productivity increased. What caused these individuals to produce at such a high rate? The answer was simple: Because these workers knew they were research participants and that they were being observed, their productivity in-

Hawthorne effect Another name for reactance or reactivity effect. creased. Thus, the knowledge that one is participating in an experiment and is being observed may result in dramatic changes in behavior. Because of the location of the original study, this reactivity phenomenon is often referred to as the **Hawthorne effect**. Having considered the general nature of naturalistic observation, we will now examine a specific observational project more closely.

Anastasia Gibson and Kristie Smith, students at the University of Alabama in Huntsville, and their faculty advisor, Aurora Torres, conducted an interesting study that used naturalistic observation. They wanted to examine the relation between the glancing behavior of people using an automated teller machine (ATM) and the proximity of other customers. On the basis of previous studies of social distance and personal space, they predicted that other potential customers would not invade the personal space of the person using the ATM and that glancing behavior would decrease as other customers approached the 4-ft. radius that defined the personal-space boundary. The researchers were "in inconspicuous locations where [they] could view the ATM" (Gibson, Smith, & Torres, 2000, p. 150). Contrary to their predictions, the results indicated that glancing behavior *increased* as proximity between the person using the ATM and other potential customers *decreased*.

As you may have surmised by now, the main drawback with the use of naturalistic observation is, once again, the inability to make cause-and-effect statements. Because we do not manipulate any variables when we use this technique, such conclusions are not possible.

Why use naturalistic observation if it does not allow us to make cause-and-effect statements? The first reason is quite straightforward: Naturalistic observation may be our only choice of research techniques to study a particular type of behavior. Psychologists who are interested in reactions to natural disasters, such as hurricanes, earthquakes, tornadoes, and fires, cannot ethically create such life-threatening situations just to study behavior; they must make their observations under naturally occurring conditions. Conducting an experiment in which researchers manipulate variables is only one of many legitimate techniques used to gather data.

A second reason for using naturalistic observation is as an adjunct to the experimental method. For example, you might use naturalistic observation before conducting an experiment to get an idea of the relevant variables involved in the situation. Once you have an idea about which variables are (and are not) important, you can conduct systematic, controlled studies of these variables in the laboratory setting. After you have conducted laboratory experiments, you may want to return to the natural setting to see whether the insights gained in the laboratory are indeed mirrored in real life. Hence, psychologists may use naturalistic observation before and after an experimental research project to acquire further information concerning relevant variables.

As with the other observation techniques, the ability to make cause-and-effect statements is the second problem with participant-as-observer research. Even though the participant observer may be close to the source of relevant information, no attempts are made to manipulate IVs or control extraneous variables.

Choosing Behaviors and Recording Techniques

It is one thing to say you are going to conduct an observational study, but it is another task actually to complete such a project. Just because the researcher does not manipulate the variables does not mean that a great deal of planning has not taken place. As is true with *any* research project, you must make several important decisions before beginning the project. Let's examine several of these decisions.

It seems simple enough to indicate that all behaviors will be observed; hence, everything of interest will be captured. Saying may be quite different from (and much easier than) doing. For example, observing and recording all behaviors may necessitate using video equipment, which may, in turn, make the observer identifiable. A participant observer with a video camera would probably not be especially effective. Had Anastasia Gibson and her colleagues (2000) used video equipment in their naturalistic study of glancing behavior at the ATM, their

obvious presence could have influenced the behavior of their participants. Hence, they remained as inconspicuous as possible when they observed the customers.

Time sampling Making observations at different time periods.

Instead of observing at exactly the same time each afternoon, the ATM researchers used a procedure known as time sampling. **Time sampling** involves making observations at different time periods in order to obtain a more representative sampling of the behavior of interest. The selection of time periods may be determined randomly or in a more systematic manner. Moreover, the use of time

sampling may apply to the same or different participants. If you are observing a group of preschoolers, using the time-sampling technique will allow you to describe the behavior of interest over a wide range of times in the same children. However, using time sampling may purposely result in the observation of different participants and increase the generality of your observations. On the other hand, Gibson et al. did not make any nighttime observations; consequently, their results may not apply to *all* times of day. In fact, they indicated that "[i]t would be interesting to see whether or not proximity or glancing behavior are influenced by the amount of sunlight at hand" (Gibson et al., 2000, p. 151).

It also is important to note that Gibson et al. (2000) observed customers at four different ATMs: Two machines were located outside ("one outside a mall and one in an urban shopping area," p. 150), whereas "two indoor ATMs were inside discount department stores" (p. 150). Why did they make observations in these different locations? Had they limited their observations to one ATM in one location, they would not be able to generalize their results beyond

that one machine to other ATMs. By observing four different ATMs, they used a technique known as situation sampling. **Situation sampling** involves observing the same type of behavior in several different situations. This technique offers the researcher two advantages. First, by sampling behavior in several different situations, you are able to determine whether the behavior in question changes as a function of the context in which you observed it. For example, a researcher might

use situation sampling to determine whether the amount of personal space people prefer differs from one culture to another or from one geographic area of a country to another.

The second advantage of the situation-sampling technique involves the fact that researchers are likely to observe different participants in the different situations. Because different individuals are observed, our ability to generalize any behavioral consistencies we notice across the various situations is increased. Because Gibson et al. (2000) made observations at four different ATMs and obtained the same results at each one, they could not attribute their findings to the customers at one specific machine.

Even if you have decided to time sample or situation sample, there is still another major decision you must make before you actually conduct your research project. You need to decide whether to present the results in a *qualitative* or *quantitative* manner. If you choose the qualitative approach, your report will consist of a description of the behavior in question (a *narrative record*) and the conclusions prompted by this description. Such narrative records can be in the form of written or tape-recorded notes that you make during or immediately after observing the behavior. Video recordings also are frequently made. If you write or tape record notes after the behavior has occurred, you should write or record them as soon as possible. In all narrative records, the language and terms used should be as clear and precise as possible, and the observer should avoid making speculative comments.

If your research plans call for a quantitative or numerical approach, you will need to know how you are going to measure the behavior under investigation and how you will

Situation sampling Observing the same behavior in different situations. analyze these measurements. We will have more to say about measurement and analysis in Chapter 9.

Using More Than One Observer: Interobserver Reliability Another consideration we must deal with in the case of observational research is whether we will use one or more observers. As the good detective knows, there are two main reasons for using more than one observer: First, one observer may miss or overlook a bit of behavior, and second, there may be some disagreement concerning exactly what was seen and how it should be rated or categorized. More than one observer may be needed even when videotape is used to preserve the complete behavioral sequence; someone has to watch the videotape and rate or categorize the behavior contained there.

When two individuals observe the same behavior, it is possible to see how well their

observations agree. The extent to which the observers agree is called **interobserver reliability**. Low interobserver reliability indicates that the observers disagree about the behavior(s) they observed; high interobserver reliability indicates agreement. Such factors as fatigue, boredom, emotional and physical state, and experience can influence

Interobserver reliability The extent to which observers agree.

interobserver reliability. If both observers are well rested, interested in their task, and in good physical and emotional health, high interobserver reliability should be obtained. An observer's physical, emotional, and attitudinal state can be monitored and dealt with easily. Additionally, the need for training observers should be considered. The importance of thorough training, especially when observing complex and subtle behaviors, cannot be stressed too much. Such training should include clear, precise definitions of the behavior(s) to be observed. The trainer should provide concrete examples of positive and negative instances of the behavior in question—if at all possible.

Even if you follow all of these guidelines, it may be difficult to obtain high interobserver reliability. In some instances the problem may reside in the type of behavior being observed and how the observers code the behaviors. For example, think of how difficult it is for two observers to agree on (a) instances of "empathy," "shame," and "self-consciousness" in 2-year-old children, (b) the difference between aggression and teasing in 10-year-olds, and (c) whether a 6-week-old infant has smiled. These examples point to the difficulty in obtaining, and the importance of having, interobserver reliability.

How can you measure interobserver reliability? A simple technique involves determining the number of agreements between the two observers and the number of opportunities the observers had to agree. Once these numbers have been determined, they are used in the following formula:

 $\frac{\text{number of times observers agree}}{\text{number of opportunities to agree}} \times 100 = \text{percentage of agreement}$

The final calculation indicates the percentage of agreement.

Another method for obtaining interobserver reliability is to calculate the correlation (see Chapter 9) between the raters' judgments and then square the correlation and multiply by 100. The resulting figure tells us the percentage of variation that is due to observer agreement; the higher the percentage, the greater the agreement.

What is a good measure of interobserver reliability? Although there are no rules to follow, a review of articles published in the two most recent issues of several journals indicated that

all articles reporting interobserver reliability had at least 85% agreement. This figure is a guideline that indicates what journal editors and reviewers consider an acceptable minimum level for interobserver reliability.

Correlational Studies

Correlational study

Determination of the relation between two variables.

relaables. In its basic form a **correlational study** involves the measurement and determination of the relation between two variables (hence the term *co-relation*). In terms of control, empirical measurement, and statistical analysis, a correlational study is likely to be more rigorous than one of the descriptive methods we've just considered. In order to understand the intent and purpose of a correlational study, we need to review some basic facts about correlations.

The Nature of Correlations

Positive correlation As scores on one variable increase, scores on the second variable also increase.

Negative correlation

As scores on one variable increase, scores on the second variable decrease.

One of three basic patterns may emerge when a correlation is calculated. The two variables may be **positively correlated**: As one variable *increases*, scores on the other variable also *increase*. For example, test scores are positively correlated if a student who makes a low score on Test 1 also scores low on Test 2, whereas a student who scores high on Test 1 also scores high on Test 2. Likewise, height and weight are positively related; in general, the taller a person is, the more he or she weighs.

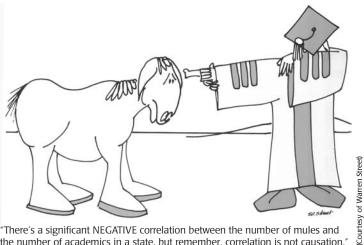
Two variables may also be negatively related. A **negative correlation** indicates that an *increase* in one variable is accompanied by a *decrease* in the second variable. For example, drinking water on a hot day and thirst are negatively related; the more water consumed, the less intense the thirst. Likewise, increasing selfesteem scores might be accompanied by decreasing anxiety scores. Thus, an

individual who scores high on a self-esteem scale would have a low score on an anxiety scale, whereas an individual who scores low on the self-esteem scale would score high on the anxiety scale.

Researchers use correlations to make predictions. For example, you probably took an entrance examination, such as the American College Test (ACT) or the Scholastic Aptitude Test (SAT), when you applied for admission to college. Previous research has shown that scores on such entrance examinations are *positively correlated* with first-semester grades in college. Thus, a college or university admissions committee might use your entrance exam score to predict how you would perform in your college classes. Obviously, the closer a correlation comes to being perfect, the better our predictions will be.



We have seen that correlations can be positive or negative in value. What has happened when we obtain a zero correlation?



"There's a significant NEGATIVE correlation between the number of mules and the number of academics in a state, but remember, correlation is not causation." (Courtesy of Warren Street.)

Zero correlations indicate that the two variables under consideration are not related. High scores on one variable may be paired with low, high, or intermediate scores on the second variable, and vice versa. In other words, knowing the score on Variable 1 does not help us predict the Zero correlation Two variables under consideration are not related.

score on Variable 2. A zero correlation may not be exactly 0. For example, a correlation of 0.03 would be considered a zero correlation by most researchers. It is important to remember that a correlation tells us only the extent to which two variables are related. Because these two variables are not under our direct control, we are still unable to make cause-and-effect statements. In short, correlations do not imply causation. It is possible that a third variable is involved. A rather farfetched example illustrates this point quite well. In an introductory statistics class, a student told one of your authors about a correlational study that reported the relation between the number of telephone poles erected in Australia each year for the 10 years following World War II and the yearly birthrate in the United States during the same time period. The result was a very high, positive correlation. The point of the example was to illustrate that correlation does not imply causation. It also illustrates the likely presence of a third variable. It is arguable that the increasingly better worldwide economic conditions that followed World War II encouraged both industrial development (the increase in telephone poles in Australia) and the desire to have a family (the higher birthrate in the United States).

Correlational Research

We agree that it is confusing when a type of mathematical analysis and a research technique have the same name. Because the common goal is to determine the relation between two variables or factors, more than coincidence is at work. Also, you should keep in mind that a correlational research project uses the correlational mathematical technique to analyze the data gathered by the researcher. Now that we have reviewed the nature and types of correlations, let's examine an example of correlational research. Kristen Robinson, a student at John Carroll University (University Heights, OH) was interested in determining what factors were good "predictors of health in order to gain knowledge of what steps can be taken to insure [sic] health during the college years" (Robinson, 2005, p. 3). To accomplish her objective, she

administered several surveys to 60 college students (30 women and 30 men). She found that locus of control and overall college adjustment were significantly correlated with illness severity. More specifically, her data showed "that students with an internal locus of control and more complete college adjustment had less severe episodes of illness than did those with an external locus of control and incomplete college adjustment" (Robinson, 2005, p. 6).

Researchers use correlational studies when data on two variables are available but they can only measure, rather than manipulate, either variable. For example, Jessica Serrano-Rodriguez, Sara Brunolli, and Lisa Echolds (2007), students at the University of San Diego, examined the relation between religious attitudes and approval of organ donation. Likewise, Adam Torres, Christy Zenner, Daina Benson, Sarah Harris, and Tim Koberlein (2007), students at Boise State University (Boise, ID), calculated the correlations between self-esteem and factors such as perceived academic abilities, family supportiveness, writing skills, and so forth, and college attendance. Although correlational investigations such as these can determine the degree of relation that exists between two variables, they are not able to offer a cause-and-effect statement concerning these variables.

REVIEW SUMMARY

- 1. **Descriptive research methods** are nonexperimental procedures for acquiring information. They do not involve the manipulation of variables.
- Some researchers make use of archival and previously recorded sources of data. Although the use of such data avoids biasing participants' responses, this approach suffers from lack of generalizability, selective deposit, and selective survival.
- **3.** Although observational techniques do not involve the direct manipulation of variables, they do allow the researcher to observe the behavior(s) of interest directly.
- 4. Naturalistic observation involves directly observing behaviors in the natural environment. In these studies the observer should remain unobtrusive in order to avoid a reactance or reactivity effect on the part of the participants.
- Because the observer may not be able to observe all behaviors at all times, decisions concerning which behaviors to observe, as well as when and where to observe them, must be made.
- Time sampling involves making observations at different time periods, whereas situation sampling involves observing the same behavior(s) in several different situations.
- 7. The use of more than one observer may be desirable to avoid missing important observations and to help resolve disagreements concerning what was or was not observed. Interobserver reliability refers to the degree to which the observers agree.
- A correlational study involves the measurement and determination of the relation between two variables.
- **9.** Two variables are **positively correlated** when an increase in one variable is accompanied by an increase in the other variable.
- Two variables are **negatively correlated** when an increase in one variable is accompanied by a decrease in the other variable.
- A zero correlation exists when a change in one variable is unrelated to changes in the second variable.

Check Your Progress

- 1. What is the reactance effect? How is it avoided by the use of archival sources of data?
- 2. Selective deposit is a problem associated with
 - a. case studies c. cause-and-effect research
 - b. naturalistic observation d. archival research
- 3. You want to write a history of the psychology department at your school, so you look through old catalogs, departmental correspondence, and musty file folders you find in the basement. You are conducting
 - a. archival research c. an experiment
 - b. a case study d. a participant observation study
- The extent to which experimenters agree about a particular bit of data best describes
 - a. archival symmetry c. concordance of naturalistic observation
 - d. participant-observer ratio b. interobserver reliability
- 5. Why are time sampling and situation sampling employed?
- 6. What is interobserver reliability? How is it calculated?
- 7. Which of the following would probably be *negatively* correlated?
 - a. number of days present in class and exam grades
 - b. height and weight of college students
 - c. number of difficult classes taken and semester GPA
 - d. overall college GPA and graduate school GPA

Ex Post Facto Studies

Can we conduct experimental research on variables that we cannot control or manipulate? Yes, we can conduct research on such variables, but we must be cautious in drawing conclu-

sions. When we work with independent variables (IVs) that we cannot or do not manipulate, we are conducting an **ex post facto study**. Ex post facto is a Latin phrase meaning "after the fact." When we conduct an ex post facto study, we are using an IV "after the fact"—it has already varied before we arrived on the scene. A great deal of detective work would seem to fall into this category. Because the experimenter has no control

Ex post facto study A study in which the variable(s) to be studied are selected after they have occurred.

over administering the IV, let alone determining who receives this variable and under what conditions it is administered, the expost facto study clearly qualifies as a descriptive research technique. However, it does have some properties in common with experimental methods.

Let's look at an example of student research using an ex post facto approach. Carolyn Licht of Marymount Manhattan College in New York City evaluated occupational stress as a function of the sex of the participants and whether the participants worked for a nonprofit or for-profit organization. Her results showed "that employees perceive more occupational stress in nonprofit than in for-profit organizations" and that "men report more stress than women in most situations" (Licht, 2000, p. 46).



What causes Licht's research to fall into the ex post facto category?

Because Licht had no control over the sex of the participants or the type of organization the participants worked for, these variables put this project in the ex post facto category. In another ex post facto study, Christy Zenner (a student at Boise State University) and her faculty supervisor, Mary Pritchard, were interested in what college students know about breast cancer and eating disorders. They hypothesized that female students would have greater knowledge of these topics than male students. Their data supported these predictions (Zenner & Pritchard, 2007). In the next section, we will consider research using surveys, questionnaires, tests, and inventories; such research is quite popular.

Surveys, Questionnaires, Tests, and Inventories

Whether they are used in experimental or nonexperimental research, investigators frequently use surveys, questionnaires, tests, and inventories to assess attitudes, thoughts, and emotions or feelings. One reason surveys and questionnaires are popular is that they appear to be quite simple to conduct; when you want to know how a group of individuals feel about a particular issue, all you have to do is ask them or give them a test. As we will see, appearances can be quite deceptive; it is not as easy to use this technique as it appears. What type of instrument is best suited for our particular project? There are numerous choices available to us. We first consider surveys and questionnaires, then turn our attention to tests and inventories.

Surveys and Questionnaires

Surveys typically request our opinion on some topic or issue that is of interest to the researcher. There are two basic types of surveys: descriptive and analytic. Although we will discuss these types separately, you should keep in mind that some surveys can serve both purposes.

Types of Surveys The purpose of your research project will determine the type of survey you choose to administer. If you seek to determine what percentage of the population has a

Descriptive survey

Seeks to determine the percentage of the population that has a certain characteristic, holds a particular opinion, or engages in a particular behavior. certain characteristic, holds a certain opinion, or engages in a particular behavior, then you will use a **descriptive survey**. The Gallup Polls that evaluate voter preferences and the Nielsen television ratings are examples of descriptive surveys. When a researcher uses this type of survey, there is no attempt to determine what the relevant variables are and how they may relate to the behavior in question. The end product is the description of a particular characteristic or behavior of a sample with the hope that this finding is representative of the population from which the sample was drawn.



The **analytic survey** seeks to determine what the relevant variables are and how they might be related. For example, Amanda Gray, a student at Agnes Scott College in Decatur, Georgia, and her faculty supervisor, Jennifer Lucas, used the Subjective Impedence Scale (Novaco, Stokols, & Milanesi, 1990) to determine "automobile commuters' subjec-

Analytic survey Seeks to determine the relevant variables and how they are related.

tive perceptions of travel impedence and their driver stress" (Gray & Lucas, 2001, p. 79). They found that one of the key determinants of commuters' stress level is the *perception* of impedence, regardless of whether it actually occurred.

What about conducting research in areas in which there are no readily available surveys? To address the issues you are interested in, you may have to construct your own survey. We *strongly* encourage you to exhaust every possibility before you attempt to construct your own survey. They look like they are easy to construct, but nothing could be further from the truth.

If you have decided that the only option is to construct a survey, you will have to choose the questions for your analytic survey *very carefully* before the survey is administered. In fact, it will

probably be necessary to do some pilot testing of the analytic survey before you use it in a full-scale investigation. **Pilot testing** refers to the testing and evaluating that is done in advance of the complete research project. During this preliminary stage the researcher tests a small number of participants and may even use in-depth interviews to help determine the type of questions that should appear on the final survey instrument.

Pilot testing Preliminary, exploratory testing that is done prior to the complete research project.

Developing a Good Survey or Questionnaire A good survey or questionnaire, one that measures the attitudes and opinions of interest in an unbiased manner, is not developed overnight; considerable time and effort typically go into its construction. When you have decided exactly what information your research project seeks to ascertain, you should follow several steps in order to develop a good survey or questionnaire. These steps appear in Table 4-1.

TABLE 4-1		Steps in Developing a Good Survey or Questionnaire	
Step 1	Decide v	what type of instrument to use. How will the information be gathered?	
Step 2	Identify the types of questions to use.		
Step 3	Write the items: They should be clear, short, and specific.		
Step 4	Pilot test and seek opinions from knowledgeable others.		
Step 5	Determine the relevant demographic data to be collected.		
Step 6	Determine administration procedures and develop instructions.		

The first step is to determine how you will obtain the information you seek. Will you use a mail survey? Will your project involve the use of a questionnaire that is administered during a regular class session at your college or university? Will trained interviewers administer your questionnaire in person, or will a telephone interview be conducted? Decisions such as these will have a major impact on the type of survey or questionnaire you will develop.

After you have decided which type of instrument you will develop, in the second step you can give attention to the *nature* of the questions that you will use and the type of responses the participants will make to these questions. Among the types of questions that are frequently used in surveys and questionnaires are the following:

1. Yes-No Questions. The respondent answers yes or no to the items.

EXAMPLE:

The thought of death seldom enters my mind. (*Source:* Templer's Death Anxiety Scale; Templer, 1970)

2. Forced Alternative Questions. The respondent must select between two alternative responses.

EXAMPLE:

A. There are institutions in our society that have considerable control over me.

B. Little in this world controls me. I usually do what I decide to do.

(Source: Reid-Ware Three-Factor Locus of Control Scale; Reid & Ware, 1973)

3. *Multiple-Choice Questions.* The respondent must select the most suitable response from among several alternatives.

EXAMPLE:

Compared to the average student,

- **A.** I give much more effort.
- **B.** I give an average amount of effort.
- C. I give less effort.

(Source: Modified Jenkins Activity Scale; Krantz, Glass, & Snyder, 1974)

4. Likert-Type Scales. The individual answers a question by selecting a response alternative from a designated scale. A typical scale might be the following: (5) strongly agree, (4) agree, (3) undecided, (2) disagree, or (1) strongly disagree.

EXAMPLE:

I enjoy social gatherings just to be with people.

1	2	3	4	5
Not at all	Not very	Slightly	Fairly	Very much
characteristic				characteristic
of me				of me

(Source: Texas Social Behavior Inventory; Helmreich & Stapp, 1974)

5. Open-Ended Questions. A question is asked to which the respondent must construct his or her own answer.

EXAMPLE:

How would you summarize your chief problems in your own words? (*Source:* Mooney Problem Check List; Mooney, 1950)

Clearly, the questions you choose to use on your survey or questionnaire will directly influence the type of data you will gather and be able to analyze when your project is completed. If you choose the yes—no format, then you will be able to calculate the frequency or percentage of such responses for each question. The use of a Likert-type scale allows you to calculate an average or mean response to each question. Should you choose to use openended questions, you will have to either decide how to code or quantify the responses or establish a procedure for preparing a summary description of each participant's answers.

The third step is to write the items for your survey or questionnaire. As a general rule, these questions should be clear, short, and specific; use familiar vocabulary; and be at the reading level of the individuals you intend to test. In preparing your items, you should avoid questions that might constrain the respondents' answers. For example, you might ask your participants to rate the effectiveness of the president of the United States in dealing with "crises." Assuming that the president has dealt with several crises, it may not be clear whether the question is referring to one particular type of crisis or another. Hence, the respondents will have to interpret the item as best they can; their interpretation may not coincide with your intended meaning. Also, researchers should avoid using questions that might bias or prejudice the respondents' answers. A negatively worded question may result in a preponderance of negative answers, whereas a positively worded question may result in a preponderance of positive answers.



Consider this yes—no question: "Do you agree that wealthy professional athletes are overpaid?" What is wrong with this question, and how can it be improved?

By indicating that professional athletes are wealthy, you have created a mental set that suggests that they may be overpaid. Thus, your respondents may be biased to answer yes. Also, using the word *agree* in the question may encourage yes answers. If the researcher rewrites the question as follows, it is less likely to bias the respondents' answers: "Do you believe professional athletes are overpaid?"

The fourth step is to pilot test your survey or questionnaire. It is important to ask others, especially professionals who have expertise in your area of research interest, to review your items. They may be able to detect biases and unintended wordings that you had not considered. It will also be helpful at this preliminary stage to administer your questionnaire to several individuals and then discuss the questions with them. Often there is nothing comparable to the insights of a participant who has actually completed a testing instrument. Such insights can be invaluable as you revise your questions. In fact, you may find it necessary to pretest and revise your survey in this manner several times before developing a final draft.

Demographic data Information about participants' characteristics such as age, sex, income, and academic major.

The fifth step involves a consideration of the other relevant information that you want your participants to provide. Frequently, such information falls under the heading of **demographic data**, which may include such items as age, sex, annual income, size of community, academic major, and academic classification. For example, in Chapter 5 we will examine research on sex differences in spatial task performance conducted by Robyn Scali (Scali & Brownlow, 2001). In their report Scali and Brownlow indicated, "Because training and prior experience may have an effect on spatial ability, we assessed experience with art, math, sports, and similar tasks" (Scali & Brownlow, 2001, p. 6); clearly, they were attentive to demographic needs.

Although the need for this step may seem obvious, it is important to review these items carefully to ensure that you have not forgotten to request a vital bit of information. We cannot begin to tell you how many survey projects designed to evaluate male–female differences were less than successful because they failed to include an item that requested the participant's sex!

The final step is to specify clearly the procedures that will be followed when the survey or questionnaire is administered. If the survey is self-administering, what constitutes the printed instructions? Are they clear, concise, and easy to follow? Who will distribute and collect the informed consent forms and deal with any questions that may arise? If your survey or questionnaire is not self-administering, then you must prepare an instruction script. The wording of this set of instructions must be clear and easily understood. Whether you present these instructions in a face-to-face interview, over the telephone, or in front of a large class of students, you (or the person giving them) must thoroughly practice and rehearse them. The researcher in charge of the project must be sure that all interviewers present the instructions in the same manner on all occasions. Likewise, questions raised by the participants must be dealt with in a consistent manner.

As we saw, the final step in creating a good survey involves a determination of the administration procedures. Because these choices are crucial to the success of this type of research, we will examine the three basic options—mail surveys, personal interviews, and telephone interviews—in more detail.

Mail Surveys Most likely you have been asked to complete a survey you received in the mail. This popular technique is used to gather data on issues that range from our opinions on environmental problems to the type of food we purchase.

One advantage of sending surveys through the mail is that the researcher does not have to be present while the survey is being completed. Thus, surveys can be sent to a much larger number of participants than a single researcher could ever hope to contact in person.

Although it is possible to put a survey in the hands of a large number of respondents, there are several disadvantages associated with this research strategy. First, the researcher cannot be sure who actually completes the survey. Perhaps the intended respondent was too busy to

complete the survey and asked a family member or friend to finish it. Hence, the time and trouble spent in creating a random sample from the population of interest may be wasted.

Even if the intended respondent completes the survey, there is no guarantee that the respondent will answer the questions in the same order in which they appeared on the survey. If the order of answering questions is relevant to the project, then this drawback may be a major obstacle to the use of mail surveys.

The low return rate associated with the use of mail surveys highlights another problem. In addition to disappointment and frustration, low return rates suggest a potential bias in the researcher's sample. What types of individuals returned the surveys? How did they differ from those individuals who did not return the surveys? Were they the least (most) busy? Were they the least (most) opinionated? We really do not know, and as the response rate drops lower, the possibility of having a biased sample increases. What constitutes a good response rate to a mail survey? It is not uncommon to have a response rate of from 25% to 30% to a mail survey; response rates of 50% and higher are considered quite acceptable.



Assume that you are planning to conduct a mail survey project. You are concerned about the possibility of having a low response rate and want to do everything to ensure the return of your surveys. What can you do to increase your response rate?

Researchers have developed strategies such as these to increase the response rates of mail surveys:

- 1. The initial mailing should include a letter that clearly summarizes the nature and importance of the research project, how the respondents were selected, and the fact that all responses are confidential. You should include a prepaid envelope for the return of the completed survey.
- 2. It may be necessary to send an additional mailing to your respondents. Because the original survey may have been misplaced or lost, it is important to include a replacement. One extra mailing may not be sufficient; you may find it necessary to send two or three requests before you achieve an acceptable response rate. These extra mailings are typically sent at 2- to 3-week intervals.

Not all researchers endorse the use of mail surveys. Low response rates, incomplete surveys, and unclear answers are among the reasons that cause some researchers to use direct interviews to obtain data. These interviews may be done in person or over the telephone.

Personal Interviews When a trained interviewer administers a survey in a respondent's home, the response rate climbs dramatically. It is not uncommon to have a 90% completion rate under these circumstances. In addition to simply increasing the response rate, the trained interviewer is able to cut down on the number of unusable surveys by clarifying ambiguous questions, making sure that all questions are answered in the proper sequence, and generally assisting with any problems experienced by the respondents.

Although this technique offers some advantages when compared to the mail survey, there are drawbacks. First, the potential for considerable expenditure of time and money exists.

Time has to be devoted to the training of the interviewers. Once trained, they will have to be paid for their time on the job. Second, the fact that an individual is administering the survey introduces the possibility of interviewer bias. Some interviewers may present some questions in a more positive (or negative) manner than do other interviewers. Only careful and extensive training of all interviewers to present all items in a consistent, neutral manner can overcome this potential difficulty. Finally, the prospect of administering surveys in the home is becoming less appealing and feasible. In many instances no one is at home during the day, and an increasing number of people are not willing to sacrifice their leisure evening hours to complete a survey. Additionally, the high crime rates in many urban areas discourage face-to-face interviewing; in its place, many investigators have turned to telephone interviewing.

Telephone Interviews In addition to overcoming several of the problems associated with personal interviews and mail surveys, telephone interviewing offers several advantages. For example, the development of random-digit dialing allows researchers to establish a random sample with ease: The desired number of calls is specified, and the computer does the rest. It is noteworthy that a random sample generated in this manner will contain both listed and unlisted telephone numbers because the digits in each number are selected randomly. With over 95% of all households in the United States currently having telephones, previous concerns about creating a biased sample consisting of only households having telephones seem largely unfounded.

Computer technology also has increased the desirability of conducting telephone interviews. For example, it is now possible to enter the participant's responses directly as they are being made. Hence, the data are stored directly in the computer and are ready for analysis at any time.

Despite these apparent advantages, telephone interviews do have potential drawbacks. Even though technology has assisted telephone researchers, it also has provided an obstacle. Many households are now equipped with answering machines or caller ID that allow incoming calls to be screened or blocked. Even if the call is answered, it is easier to say no to an unseen interviewer on the telephone than to a person at your front door. Additionally, people who have added call blocking to their phone service will not be included in the sample. These three situations lower the response rate and raise the possibility of a biased sample.

The use of the telephone also prohibits the use of visual aids that might serve to clarify certain questions. In addition, because the telephone interviewer cannot see the respondent, it is not possible to evaluate nonverbal cues such as facial expressions, gestures, and posture. Such cues might suggest that a certain question was not completely understood or that an answer is in need of clarification.

Not being in face-to-face contact with the respondent also makes it more difficult to establish rapport. Hence, telephone respondents may not be as willing to participate in the survey. This potential lack of willingness has led to the use of shorter survey instruments.

Although surveys and questionnaires are popular research tools with many investigators, there are other ways to gather data. Since the late 1800s when Sir Francis Galton (1822–1911) attempted to evaluate people's ability or intelligence by measuring physical attributes such as reaction time or visual acuity, psychologists have developed a large number of tests and inventories for a wide variety of purposes.

Tests and Inventories

Unlike surveys and questionnaires, which evaluate opinions on some topic or issue, tests and inventories are designed to assess a specific attribute, ability, or characteristic possessed by

the individual being tested. In this section we look at the characteristics of good tests and inventories and then discuss three general types of tests and inventories: achievement, aptitude, and personality.

Characteristics of Good Tests and Inventories Unlike surveys and questionnaires, the researcher is less likely to be directly involved with the development of a test or inventory. Because their development and pilot testing has already taken place, you will need to scruti-

nize the reports concerning the development of each test or inventory you are considering. A good test or inventory should possess two characteristics: It should be valid, and it should be reliable.

Validity A test or inventory has **validity** when it actually measures what it is supposed to measure. If your research calls for a test that measures spelling achievement, you want the instrument you select to measure that ability, not another accomplishment, such as mathematical proficiency.

There are several ways to establish the validity of a test or inventory. **Content validity** indicates that the test items actually represent the type of material they are supposed to test. Researchers often use a panel of expert judges to assess the content validity of test items. Although the more subjective evaluation of such judges may not lend itself to a great deal of quantification, their degree of agreement, known as **interrater relia-bility**, can be calculated. Interrater reliability is similar to interobserver reliability. The main difference is that interrater reliability measures agreement between judgments concerning a test item, whereas interobserver reliability measures agreement between observations of behavior.

We can establish **concurrent validity** when we already have another measure of the desired trait or outcome and can compare the score on the test or inventory under consideration with this other measure. For example, the scores made by a group of patients on a test designed to measure aggression might be compared with a diagnosis of their aggressive tendencies made by a clinical psychologist. If the test and the clinical psychologist rate the aggressiveness of the patients in a similar manner, then concurrent validity for the test has been established.

Often our second measure may not be immediately accessible. When the test score is to be compared with an outcome that will occur in the future, the researcher is attempting to establish the **criterion validity** of the test. Thus, criterion validity refers to the ability of the test or inventory to predict the outcome or criterion. For example, it is the desired outcome that college entrance examinations such as the SAT and ACT predict first-semester performance in college. To the extent that these tests are successful at predicting first-semester GPAs, their criterion validity has been established.

Reliability Once we have determined that a particular test is valid, we also want to make sure that it is reliable. **Reliability** refers to the extent to which the test or inventory is consistent in its evaluation of the same individuals over repeated administrations. For example, if we have

Validity The extent to which a test or inventory measures what it is supposed to measure.

Content validity The extent to which test items actually represent the type of material they are supposed to represent.

Interrater reliability Degree of agreement among judges concerning the content validity of test or inventory items.

Concurrent validity

Degree to which the score on a test or inventory corresponds with another measure of the designated trait.

Criterion validity

Established by comparing the score on a test or inventory with a future score on another test or inventory.

Reliability Extent to which a test or inventory is consistent in its evaluation of the same individuals. developed a test to measure aptitude for social work, we would want individuals who score high (or low) on our test on its first administration to make essentially the same score when

they take the test again. The greater the similarity between scores produced by the same individuals on repeated administrations, the greater the reliability of the test or inventory.

Reliability is typically assessed through the test–retest or split-half procedures. When the **test–retest procedure** is used, the test is simply given a second time and the scores from the two tests are compared; the greater the similarity, the higher the reliability.



On the surface, the test–retest procedure appears to be quite straightforward and reasonable; however, there may be a problem with establishing reliability. What is the problem?

The main problem with the test–retest procedure comes from the fact that the participants are repeatedly administered the same test or inventory. Having already taken the test or inventory, the individuals may remember the questions and answers the next time the instrument is administered. Their answers may be biased by the previous administration. If a lengthy time period elapses between administrations, the participants might forget the questions and answers, and this familiarity problem might be overcome. On the other hand, a lengthy time period may influence reliability in yet another manner. An extended time period allows the participants to have numerous experiences and learning opportunities between administrations. These experiences, called *history effects* (see Chapter 8), may influence their scores when the participants take the test or inventory again. Hence, the reliability measure may be influenced by the experiences that intervene between the two testing sessions.

Split-half technique

Test-retest procedure

a test to the same

participants.

Determination of reliability

by repeatedly administering

Determination of reliability by dividing the test or inventory into two subtests and then comparing the scores made on the two halves. It is possible to overcome the problems of test familiarity and lengthy time periods separating administrations by using the split-half approach. The **split-half technique** of establishing reliability involves dividing a test or inventory into two halves or subtests and then administering them to the same individuals on different occasions or by administering the entire test and then splitting it into two halves. Because the questions that comprise the two subtests came from the same test, it is assumed that they are highly related to each other if the test is reliable. Typically, the questions that comprise these two subtests are selected randomly or in some predetermined manner, such as odd-even. The higher the degree of correspondence

between scores on the two subtests, the greater the reliability of the overall test from which they were selected.

Having determined that a test or inventory should be valid and reliable, we now examine several types of these instruments that are currently used for research and predictive purposes.

Achievement test

Designed to evaluate an individual's level of mastery or competence.

Types of Tests and Inventories Achievement tests are given when an evaluation of an individual's level of mastery or competence is desired. For example, doctors must pass a series of medical board examinations before they are allowed to practice medicine, and lawyers must pass the bar examination before they are allowed to practice law. The score that distinguishes passing from failing determines the minimum level of achievement that must be attained. You can probably think of many achievement tests you have taken during your life.

At many colleges and universities the counseling center or career development office

offers students the opportunity to take an aptitude test to assist them in selecting a major or making a career choice. An **aptitude test** is used to assess an individual's ability or skill in a particular situation or job. For example, the Purdue Pegboard Test is often administered to determine aptitude for jobs that require manual dexterity. According to Anastasi (1988), "[T]his test provides a measure of two types of activity, one re-

quiring gross movements of hands, fingers, and arms, and the other involving tip-of-thefinger dexterity needed in small assembly work" (p. 461). Similarly, if you are planning to attend graduate school, you will probably be required to take the Graduate Record Examination (GRE). For most graduate schools, the two most important scores on the GRE are the verbal and quantitative subtests. These scores represent measures of your aptitude to complete successfully verbal and quantitative courses on the graduate level.

The **personality test or inventory** measures a specific aspect of an individual's motivational state, interpersonal capability, or personality (Anastasi & Urbina, 1997). The use of a personality inventory in research is exemplified by a project reported by Dana Bodner and C. D. Cochran, students at Stetson University in DeLand, Florida, and their faculty advisor, Toni Blum (Bodner, Cochran, & Blum, 2000). The purpose of this study was to validate a scale, the General Unique Invulnerability Scale (GUI), which

measures general optimism about unique invulnerability (the belief that you will not experience negative events or misfortunes). These researchers administered the GUI to a sample of 40 skydivers and a sample of 40 college students and found that skydivers had higher GUI scores than did the college students. It is not surprising that skydivers would have a stronger belief in unique invulnerability than the typical college student who does not engage in this risky behavior.

Sampling Considerations and Basic Research Strategies

Having completed our review of surveys, questionnaires, tests, and inventories, we conclude this chapter with a consideration of two more issues that all researchers must consider when they conduct a project: sampling and basic research strategies. Sampling deals with the question of who will participate in your research project and whether the participants are a representative group. Once you have dealt with sampling issues, then you must decide on your main research strategy. The main strategies used by researchers are the single-strata, crosssectional, and longitudinal approaches.

Sampling

Assume that you want to determine which of two new titles for the college newspaper—*The Wilderbeast* (named after your school mascot) or *The Observer*—appeals most to the student body. You ask the 36 students in your senior-level biopsychology course and find that 24 prefer *The Observer* and 12 prefer *The Wilderbeast*. You report your findings to the publications advisory board and recommend that *The Observer* be chosen as the new title.

Aptitude test Designed to assess an individual's potential ability or skill in a particular job.

Personality test or inventory Measures a specific aspect of the individual's motivational state, interpersonal capability, or personality.



Should the publications board accept your recommendation? Are there some reasons to question your findings?

The publications advisory board should not accept your recommendation. Unlike Sherlock Holmes, who said, "I presume nothing" (Doyle, 1927, p. 745), you seem to have made some assumptions. The main problem with your data concerns the students you surveyed. Is your senior-level biopsychology class representative of the entire campus? The answer to this question must be a resounding no; only senior-level psychology majors take this class. Moreover, a quick check of the class roster indicates that the majority (67%) of the students in

your class are women. Clearly, you should have selected a group of students more representative of the general student body at your college.

Population The complete set of individuals or events.

Sample A group that is selected to represent the population.

Random sample A

sample in which every member of the population has an equal likelihood of being included.

Random sampling without replacement

Once chosen, a score, event, or participant cannot be returned to the population to be selected again.

Random sampling with replacement Once chosen, a score, event, or participant can be returned to the population to be selected again. We shall designate the general student body as the **population** or the complete set of individuals or events we want to represent. The group that we select to represent the population is called the **sample**. When every member of the population has an equal likelihood of being selected for inclusion in the sample, we have created a **random sample**.

How would you obtain a random sample of students on your campus to take the newspaper title survey? Computer technology has made this task quite simple; you simply indicate the size of the sample you desire, and the computer can be programmed to randomly select a sample of that size from the names of all currently enrolled students. Because a name is not eligible to be chosen again after it has been selected, this technique is called **random sampling without replacement**. If the chosen item can be returned to the population and is eligible to be selected again, the procedure is termed **random sampling with replacement**. Because psychologists do not want the same participant to appear more than once in a group, random sampling without replacement is the preferred technique for creating a random research sample.

Suppose a sample of 80 students has been randomly selected from the entire student body by the computer and you are examining the printout. Even though you selected this sample randomly, it also has some apparent problems. Just by chance the majority of the students selected are freshmen and sophomores. Moreover, the majority of students in this sample are men. The views on the two newspaper titles held by this group of randomly selected students may not be much more representative than our original sample. What can we do to produce an even more representative sample?

There are two techniques that we can use to increase the representativeness of our sample. The first procedure is quite simple: We can select a larger sample. Generally speaking, the larger the sample, the more representative it will be of the population. If we randomly selected 240 students, this larger sample would be more representative of the general student body than our original sample of 80 students.

Although larger samples may be more like the population from which they are drawn, there is a potential drawback. Larger samples mean that more participants will need to be tested. In our project dealing with the two newspaper titles, the testing of additional participants may not present any major problems. However, if we were administering a lengthy questionnaire or paying participants for their participation, increasing the number of participants might create unmanageable time or financial obstacles.

If simply increasing the sample size does not offer a good solution to the problem of

achieving a representative sample, the researcher may want to use stratified random sampling. **Stratified random sampling** involves dividing the population into subpopulations or strata and then drawing a random sample from one or more of these strata. For example, one logical subdivision of a college student body would be by classes: freshmen, sophomores, juniors, and seniors. You could then draw a random sample from each class. How many students will be included in each stratum? One option would be for each stratum to contain

Stratified random sampling Random samples are drawn from specific subpopulations or strata of the general population.

an equal number of participants. Thus, in our newspaper title project we might include 20 participants from each academic classification. A second option would be to sample each stratum in proportion to its representation in the population. If freshmen comprise 30% of the student body, then our random sample would contain 30% first-year students. What about the number of men and women sampled in each stratum? We could have equal numbers, or we could sample in proportion to the percentage of men and women in each stratum. As you have probably surmised, the use of stratified random sampling indicates that you have considerable knowledge of the population in which you are interested. Once you have this knowledge, you can create a sample that is quite representative of the population of interest. A word of caution is in order, however. Although it may be tempting to specify a number of characteristics that your sample must possess, you can carry this process too far. If your sample becomes too highly specified, then you will be able to generalize or extend your results only to a population having those very specific characteristics.

One stratum that frequently appears in research reports is the subject or participant pool, which is used by many colleges and universities. Here, students, typically enrolled in introductory psychology, have the option (among others) of participating in psychological research in order to fulfill a course requirement. After students have volunteered to participate in a research project, the researcher can randomly assign them to specific groups or treatment conditions. For example, in the research on the wording of the informed consent document done by Burkley et al. (2000), which we considered in Chapter 2, the authors stated,

Twenty-five undergraduate university psychology students (2 men, 23 women) volunteered to participate. Participants were randomly assigned to either the control or experimental group. Participants received class credit for their involvement in the study. (p. 44)

Basic Research Strategies

Even though you have chosen your sample, you cannot simply rush out to start testing participants. You need to give some thought to the research question and how you can best conduct your project in order to answer that question. There are three basic approaches you can adopt: single-strata, cross-sectional, and longitudinal.

Single-strata approach

Gathering data from a single stratum of the population of interest.

Cross-sectional

research Comparison of two or more groups during the same, rather limited, time period.

Longitudinal research

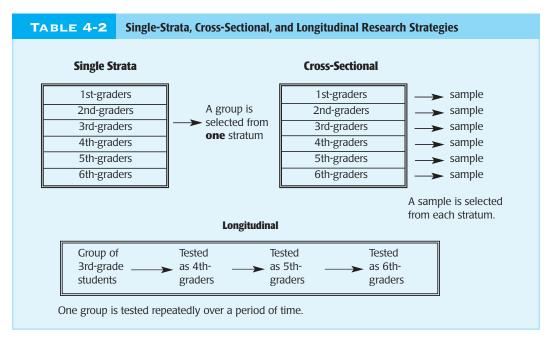
project Obtaining research data from the same group of participants over an extended period of time.

Cohort A group of individuals born during the same time period.

The **single-strata approach** seeks to acquire data from a single, specified segment of the population. For example, a particular Gallup Poll may be interested only in the voting preferences of blue-collar workers. Hence, a sample composed only of individuals from this stratum would be administered a voter-preference survey. This approach typically seeks to answer a rather specific research question.

When the single-strata approach is broadened to include samples from more than one stratum, a cross-sectional approach is being employed. **Cross-sectional research** involves the comparison of two or more groups of participants during the same, rather limited, time span. For example, a researcher may want to compare voter preferences of different age groups. To acquire this information, random samples of voters ages 21, 31, 41, 51, 61, and 71 are obtained and their responses to a voter-preference survey are compared.

Perhaps the researcher wants to obtain information from a group of participants over an extended period of time. In this instance a **longitudinal research project** would be conducted. First, the researcher would obtain a random sample from the population of interest; then this sample would complete an initial survey or test. The same participants would then be contacted periodically to determine what, if any, changes had occurred during the ensuing time in the behavior of interest. This group of individuals, born in the same time period and repeatedly surveyed or tested, is called a **cohort**. For example, a researcher might be interested in changes in the degree of support for environmental conservation that may occur as individuals grow older. To evaluate such changes, a group of grade-school children is randomly selected. Every five years, all members of this cohort are contacted and administered an environmental conservation survey. Table 4-2 allows you to compare these three different research strategies.



It is important that you remember that sampling concerns and decisions about your basic research strategy apply to both nonexperimental and experimental research projects. Moreover, our concerns with sampling and research strategy do not end at this point. For example, we will have more to say about the importance of randomization when we discuss experimental control in Chapter 6. Likewise, once we have decided on our basic research strategy, we will have to consider such details as how many groups of participants must be tested to answer our research question. These topics fall under the general heading of experimental design; we begin our discussion of this topic in Chapter 10.

REVIEW SUMMARY

- 1. In an **ex post facto study** the variables have been experienced before they are examined by the researcher; therefore, control and manipulation of variables cannot be accomplished.
- Surveys, questionnaires, tests, and inventories are used to assess attitudes, thoughts, and emotions or feelings.
- Descriptive surveys seek to determine the percentage of a population that has a certain characteristic, holds a certain opinion, or engages in a particular behavior.
 Analytic surveys seek to determine the variables that are relevant in a situation and their relation.
- 4. The steps to be completed in developing a good survey or questionnaire include considering the type of instrument to be developed, determining the types of questions to be used, writing the items, **pilot testing**, determining the relevant demographic data to be collected, and deciding on administration procedures.
- Demographic data include relevant information, such as gender, age, income, and educational level, about the participants.
- 6. Mail surveys can be sent to a large number of potential respondents; however, the researcher cannot be sure who actually completed the survey or in what order they completed the questions. Low response rates for mail surveys can be improved by stressing the importance of the project and by sending additional mailings.
- **7.** Personal interviews yield a higher rate of completed surveys, but they are costly in terms of time and money. The increase in the number of working families and the escalating crime rate in urban areas have made the use of personal interviews less desirable.
- **8.** Telephone interviews allow the researcher to reach a large number of respondents more efficiently than personal interviews and mail surveys. On the other hand, answering machines and the inability to see nonverbal cues are drawbacks to this approach.
- Tests and inventories should be valid (measure what they are supposed to measure) and reliable (be consistent in their evaluation).
- 10. Validity may be established by the content, concurrent, and criterion methods.
- 11. The test-retest and split-half procedures are used to establish reliability.
- 12. Achievement tests evaluate level of mastery or competence. Aptitude tests assess an individual's ability or skill in a particular situation or job. A personality test or inventory measures a specific aspect of the individual's motivational state, interpersonal capability, or personality.

- **13.** The **sample** of individuals who complete a survey, questionnaire, test, or inventory should be representative of the **population** from which it is drawn.
- 14. When a random sample is selected, every member of the population has an equal likelihood of being selected. When random sampling without replacement is used, an item cannot be returned to the population once it has been selected. When random sampling with replacement is used, selected items can be returned to the population to be selected again.
- **15. Stratified random sampling** involves dividing the population into subpopulations or strata and then drawing random samples from these strata.
- 16. Basic research strategies include investigating (a) a single stratum of a specified population, (b) samples from more than one stratum in a cross-sectional project, or (c) a single group of participants over an extended time period in a longitudinal study.

Check Your Progress

- 1. Matching
 - 1. descriptive survey
 - 2. analytic survey
 - 3. pilot testing
 - 4. demographic data
 - 5. cohort
 - 6. concurrent validity
 - 7. criterion validity

- A. defines participants born in the same time period
- B. affects ability to predict the outcome
- C. may include age, sex, and annual income
- D. indicates the percentage having a certain characteristic
- E. compares scores on two separate measures
- F. tries to determine what the relevant variables are
- G. does testing or evaluating in advance of the complete research project
- **2.** Describe the steps involved in creating a good survey.
- 3. "Working with IVs that the experimenter does not manipulate" best describes
 - a. a case study
 - b. naturalistic observation
 - c. participant observation
 - d. an ex post facto study
- 4. How can the low return rate of mail surveys be improved?
- 5. Why is the use of personal interviews declining?
- 6. Distinguish between achievement and aptitude tests.
- **7.** _____ are used to assess a specific attribute or ability.
 - a. Surveys
 - b. Questionnaires
 - c. Pilot studies
 - d. Inventories

- 8. The general group of interest is the _____. The group that is selected to represent the general group is a _____.
- 9. A test can be _____ and not be _____.
 - a. valid; accurate
 - b. reliable; valid
 - c. valid; reliable
 - d. split; halved
- **10.** What is random sampling? With replacement? Without replacement?
- 11. What is stratified random sampling, and why is it used?
- **12.** Distinguish among the single-strata, cross-sectional, and longitudinal approaches to research.

Key Terms

Descriptive research methods, 59	Pilot testing, 71	Population, 80	
Naturalistic observation, 61	Demographic data, 74	Sample, 80	
Reactance or reactivity effect, 62	Validity, 77	Random sample, 80	
Hawthorne effect, 62	Content validity, 77	Random sampling without replacement, 80	
Time sampling, 64	Interrater reliability, 77		
Situation sampling, 64	Concurrent validity, 77	Random sampling with	
Interobserver reliability, 65	Criterion validity, 77	replacement, 80 Stratified random sampling, 81	
Correlational study, 66	Reliability, 77		
Positive correlation, 66	Test-retest procedure, 78		
Negative correlation, 66	Split-half technique, 78	Single-strata approach, 82	
Zero correlation, 67	Achievement test, 78	Cross-sectional research, 82	
Ex post facto study, 69	Aptitude test, 79	Longitudinal research	
Descriptive survey, 70	Personality test or inventory, 79	project, 82 Cohort, 82	
Analytic survey, 71			

Looking Ahead

In Chapters 3 and Chapter 4 we have considered approaches to gathering data that do not include the direct manipulation of any variables or factors by the researcher. These approaches therefore do not qualify as true experiments. In Chapter 5 we begin our consideration of experiments. First we carefully examine the scientific method and the variables involved in an experiment; then we consider the procedures involved in the control of these variables.