Chapter 6

Descriptive Research

Our physical and social worlds present overwhelming amounts of information. But if you study a well-chosen sample from one of those worlds—and draw reasonable inferences from your observations of this sample—you can learn a great deal.

Learning Outcomes

6.1 Describe general characteristics and purposes of (a) observation studies, (b) correlational research, (c) developmental designs, and (d) survey research. Also, describe effective strategies you might use in each of these four research methodologies.

6.2 Identify effective strategies for conducting a face-to-face, telephone, or video-conferencing interview.

6.3 Identify effective strategies for constructing and administering a questionnaire and for analyzing people's responses to it.

6.4 Explain possible uses of checklists, rating scales, rubrics, computer software, and the Internet in data collection.

6.5 Determine an appropriate sample for a descriptive study.

6.6 Describe common sources of bias in descriptive research, as well as strategies for minimizing the influences of such biases.

In this chapter, we discuss types of quantitative study that fall under the broad heading *descriptive quantitative research.* This general category of research designs involves either identifying the characteristics of an observed phenomenon or exploring possible associations among two or more phenomena. In every case, **descriptive research** examines a situation *as it is.* It does not involve changing or modifying the situation under investigation, nor is it intended to determine cause-and-effect relationships.

DESCRIPTIVE RESEARCH DESIGNS

In the next few pages, we describe observation studies, correlational research, developmental designs, and survey research, all of which yield quantitative information that can be summarized through statistical analyses. We devote a significant portion of the chapter to survey research, because this approach is used quite frequently in such diverse disciplines as business, government, public health, sociology, and education.

Observation Studies

As you will discover in Chapter 9, many qualitative researchers rely heavily on personal observations—typically of people or another animal species (e.g., gorillas, chimpanzees)—as a source of data. In *quantitative* research, however, an **observation study** is quite different. For one thing, an observation study in quantitative research might be conducted with plants rather than animals, or it might involve nonliving objects (e.g., rock formations, soil samples) or dynamic physical phenomena (e.g., weather patterns, black holes).

Also, a quantitative observation study tends to have a limited, prespecified focus. When human beings are the topic of study, the focus is usually on a certain aspect of behavior. Furthermore, the behavior is quantified in some way. In some situations, each occurrence of the behavior is *counted* to determine its overall frequency. In other situations, the behavior is *rated* for accuracy, intensity, maturity, or some other dimension. But regardless of approach, the researcher strives to be *as objective as possible* in assessing the behavior being studied. To maintain such objectivity, he or she is likely to use strategies such as the following:

- Define the behavior being studied in such a precise, concrete manner that the behavior is easily recognized when it occurs.
- Divide the observation period into small segments and then record whether the behavior does or does not occur during each segment. (Each segment might be 30 seconds, 5 minutes, 15 minutes, or whatever other time span is suitable for the behavior being observed.)
- Use a rating scale to evaluate the behavior in terms of specific dimensions (more about rating scales later in the chapter).
- Have two or three people rate the same behavior independently, without knowledge of one another's ratings.
- Train the rater(s) to use specific criteria when counting or evaluating the behavior, and continue training until consistent ratings are obtained for any single occurrence of the behavior.

A study by Kontos (1999) provides an example of what a researcher might do in an observation study. Kontos's research question was this: What roles do preschool teachers adopt during children's free-play periods? (She asked the question within the context of theoretical issues that are irrelevant to our purposes here.) The study took place during free-play sessions in Head Start classrooms, where 40 preschool teachers wore cordless microphones that transmitted what they said (and also what people near them said) to a remote audiotape recorder. Each teacher was audiotaped for 15 minutes on each of two different days. Following data collection, the tapes were transcribed and broken into 1-minute segments. Each segment was coded in terms of the primary role the teacher assumed during that time, with five possible roles being identified: *interviewer* (talking with children about issues unrelated to a play activity), *stage manager* (helping children get ready to engage in a play activity), *play enhancer/playmate* (joining a play activity in some way), *safety/behavior monitor* (managing children's behavior), or *uninvolved* (not attending to the children's activities in any manner). Two research assistants were trained in using this coding scheme until they were consistent in their judgments at least 90% of the time, indicating a reasonably high *interrater reliability.* They then independently coded each of the 1-minute segments and discussed any segments on which they disagreed, eventually reaching consensus on all segments. (The researcher found, among other things, that teachers' behaviors were to some degree a function of the activities in which the children were engaging. Her conclusions, like her consideration of theoretical issues, go beyond the scope of this book.)

As should be clear from the preceding example, an observation study involves considerable advance planning, meticulous attention to detail, a great deal of time, and, often, the help of one or more research assistants. Furthermore, a pilot study is essential for ironing out any wrinkles in identifying and classifying the behavior(s) or other characteristic(s) under investigation. Embarking on a full-fledged study without first pilot testing the methodology can result in many hours of wasted time.

Ultimately, an observation study can yield data that portray some of the richness and complexity of human behavior. In certain situations, then, it provides a quantitative alternative to such qualitative approaches as ethnographies and grounded theory studies (see Chapter 9).

Correlational Research

A **correlational study** examines the extent to which differences in one characteristic or variable are associated with differences in one or more *other* characteristics or variables. A **correlation** exists if, when one variable increases, another variable either increases or decreases in a somewhat data were to include only children of ages 6 and 7, we would have greater homogeneity with respect to reading ability than would be the case for a sample of children ages 6 through 13. Second, we can *describe* the degree to which the two variables are intercorrelated, perhaps by computing a statistic known as a *correlation coefficient* (Chapter 8 provides details). But third—and most importantly—we can *interpret* these data and give them meaning. The data tell us not only that children become better readers as they grow older—that's a "no brainer"—but also that any predictions of children's future reading abilities based on age alone will be imprecise ones at best.

A Caution About Interpreting Correlational Results

When two variables are correlated, researchers sometimes conclude that one of the variables must in some way cause or influence the other. In some instances, such an influence may indeed exist; for example, chronological age—or at least the amount of experience that one's age reflects—almost certainly has a direct bearing on children's mental development, including their reading ability. But ultimately we can never infer a cause-and-effect relationship on the basis of correlation alone. Simply put, *correlation does not, in and of itself, indicate causation.*

Let's take a silly example. A joke that seems to have "gone viral" on the Internet is this one:

> I don't trust joggers. They're always the ones that find the dead bodies. I'm no detective . . . just sayin'.

The tongue-in-cheek implication here is that people who jog a lot are more likely to be murderers than people who don't jog very much and that perhaps jogging *causes* someone to become a murderer—a ridiculous conclusion! The faulty conclusion regarding a possible cause-and-effect relationship is crystal clear.

In other cases, however, it would be all too easy to draw an unwarranted cause-and-effect conclusion on the basis of correlation alone. For example, in a series of studies recently published in the journal *Psychological Science,* researchers reported several correlations between parenthood and psychological well-being: Adults who have children tend to be happier—and to find more meaning in life—than adults who don't have children (Nelson, Kushlev, English, Dunn, & Lyubomirsky, 2013). Does this mean that becoming a parent *causes* greater psychological well-being? Not necessarily. Possibly the reverse is true—that happier people are more likely to *want* to have children, and so they take steps to have them either biologically or through adoption. Or perhaps some other factor is at the root of the relationship—maybe financial stability, a strong social support network, a desire to have a positive impact on the next generation, or some other variable we haven't considered.

The data may not lie, but the causal conclusions we draw from the data may, at times, be highly suspect. Ideally, a good researcher isn't content to stop at a correlational relationship, because *beneath the correlation* may lie some potentially interesting dynamics. One way to explore these dynamics is through *structural equation modeling (SEM),* a statistical procedure we describe briefly in Table 8.5 in Chapter 8. Another approach—one that can yield more solid conclusions about cause-and-effect relationships—is to follow up a correlational study with one or more of the experimental studies described in Chapter 7 to test various hypotheses about what causes what.

Developmental Designs

Earlier we presented a hypothetical example of how children's ages might correlate with their reading levels. Oftentimes when researchers want to study how a particular characteristic changes as people grow older, they use one of two developmental designs, either a cross-sectional study or a longitudinal study.

In a **cross-sectional study**, people from several different age-groups are sampled and compared. For instance, a developmental psychologist might study the nature of friendships for children at ages 4, 8, 12, and 16. A gerontologist might investigate how retired people in their 70s, 80s, and 90s tend to spend their leisure time.

To address some of the weaknesses of longitudinal and cross-sectional designs, researchers occasionally combine both approaches in what is known as a **cohort-sequential study**. In particular, a researcher begins with two or more age-groups (this is the cross-sectional piece) and follows each age-group over a period of time (this is the longitudinal piece). As an example, let's return to the issue of how people's logical thinking ability changes over time. Imagine that instead of doing a simple cross-sectional study involving 20-year-olds and 70-year-olds, we begin with a group of 20-year-olds and a group of 65-year-olds. At the beginning of the study, we give both groups a multiple-choice test designed to assess logical reasoning; then, 5 years later, we give the test a second time. If both groups improve over the 5-year time span, we might wonder if practice in taking multiple-choice tests or practice in taking this *particular* test might partly account for the improvement. Alternatively, if the test scores increase for the younger (now 25-year-old) group but decrease for the older (now 70-year-old) group, we might reasonably conclude that logical thinking ability *does* decrease somewhat in the later decades of life.

Like a longitudinal study, a cohort-sequential study enables us to calculate correlations between measures taken at two different time periods and therefore to make predictions across time. For instance, we might determine whether people who score highest on the logical thinking test at Time 1 (when they are either 20 or 65 years old) are also those who score highest on the test at Time 2 (when they are either 25 or 70 years old). If we find such a correlation, we can reasonably conclude that logical thinking ability is a relatively stable characteristic—that certain people currently think and will continue to think in a more logical manner than others. We could also add other variables to the study—for instance, the amount of postsecondary education that participants have had and the frequency with which they engage in activities that require logical reasoning—and determine whether such variables mediate or moderate the long-term stability of logical reasoning ability.

Cross-sectional, longitudinal, and cohort-sequential designs are used in a variety of disciplines, but as you might guess, they are most commonly seen in developmental research (e.g., studies in child development or gerontology). Should you wish to conduct a developmental study, we urge you to browse in such journals as *Child Development* and *Developmental Psychology* for ideas about specific research strategies.

Survey Research

Some scholars use the term *survey research* to refer to almost *any* form of descriptive, quantitative research. We use a more restricted meaning here: **Survey research** involves acquiring information about one or more groups of people—perhaps about their characteristics, opinions, attitudes, or previous experiences—by asking them questions and tabulating their answers. The ultimate goal is to learn about a large population by surveying a sample of that population; thus, we might call this approach a *descriptive survey* or *normative survey.*

Reduced to its basic elements, a *survey* is quite simple in design: The researcher poses a series of questions to willing participants; summarizes their responses with percentages, frequency counts, or more sophisticated statistical indexes; and then draws inferences about a particular population from the responses of the sample. It is used with more or less sophistication in many areas of human activity—for instance, in a neighborhood petition in support of or against a proposed town ordinance or in a national telephone survey seeking to ascertain people's views about various candidates for political office. This is not to suggest, however, that because of their frequent use, surveys are any less demanding in their design requirements or any easier for the researcher to conduct than other types of research. Quite the contrary, a survey design makes critical demands on the researcher that, if not carefully addressed, can place the entire research effort in jeopardy.

Survey research captures a fleeting moment in time, much as a camera takes a single-frame photograph of an ongoing activity. By drawing conclusions from one transitory collection of data, we might generalize about the state of affairs for a longer time period. But we must keep in mind the wisdom of the Greek philosopher Heraclitus: There is nothing permanent but change.

Survey research typically employs a face-to-face interview, a telephone interview, or a written questionnaire. We discuss these techniques briefly here and then offer practical suggestions for conducting them in "Practical Application" sections later on. We describe a fourth

PLANNING FOR DATA COLLECTION IN A DESCRIPTIVE STUDY

Naturally, a descriptive quantitative study involves measuring one or more variables in some way. With this point in mind, let's return to a distinction first made in Chapter 4: the distinction between substantial and insubstantial phenomena. When studying the nature of *substantial phenomena*—phenomena that have physical substance, an obvious basis in the physical world—a researcher can often use measurement instruments that are clearly valid for their purpose. Tape measures, balance scales, oscilloscopes, MRI machines—these instruments are indisputably valid for measuring length, weight, electrical waves, and internal body structures, respectively. Some widely accepted measurement techniques also exist for studying *insubstantial phenomena*—concepts, abilities, and other intangible entities that cannot be pinned down in terms of precise physical qualities. For example, an economist might use Gross Domestic Product statistics as measures of a nation's economic growth, and a psychologist might use the *Stanford-Binet Intelligence Scales* to measure children's general cognitive ability.

Yet many descriptive studies address complex variables—perhaps people's or animals' day-to-day behaviors, or perhaps people's opinions and attitudes about a particular topic—for which no ready-made measurement instruments exist. In such instances, researchers often collect data through systematic observations, interviews, or questionnaires. In the following sections, we explore a variety of strategies related to these data-collection techniques.

PRACTICAL APPLICATION Using Checklists, Rating Scales, and Rubrics

Three techniques that can facilitate quantification of complex phenomena are checklists, rating scales, and rubrics. A **checklist** is a list of behaviors or characteristics for which a researcher is looking. The researcher—or in many studies, each participant—simply indicates whether each item on the list is observed, present, or true or, in contrast, is *not* observed, present, or true.

A **rating scale** is more useful when a behavior, attitude, or other phenomenon of interest needs to be evaluated on a continuum of, say, "inadequate" to "excellent," "never" to "always," or "strongly disapprove" to "strongly approve." Rating scales were developed by Rensis Likert in the 1930s to assess people's attitudes; accordingly, they are sometimes called **Likert scales**.[2]

Checklists and rating scales can presumably be used in research related to a wide variety of phenomena, including those involving human beings, nonhuman animals, plants, or inanimate objects (e.g., works of art and literature, geomorphological formations). We illustrate the use of both techniques with a simple example involving human participants. In the late 1970s, park rangers at Rocky Mountain National Park in Colorado were concerned about the heavy summertime traffic traveling up a narrow mountain road to Bear Lake, a popular destination for park visitors. So in the summer of 1978, they provided buses that would shuttle visitors to Bear Lake and back again. This being a radical innovation at the time, the rangers wondered about people's reactions to the buses; if there were strong objections, other solutions to the traffic problem would have to be identified for the following summer.

Park officials asked a sociologist friend of ours to address their research question: How do park visitors feel about the new bus system? The sociologist decided that the best way to approach the problem was to conduct a survey. He and his research assistants waited at the parking lot to which buses returned after their trip to Bear Lake; they randomly selected people who exited the bus and administered the survey. With such a captive audience, the response rate was extremely high: 1,246 of the 1,268 people who were approached agreed to participate in the study, yielding a response rate of 98%.

[2]Although we have often heard *Likert* pronounced as "lie-kert," Likert pronounced his name "lick-ert."

FIGURE 6.4 ■ Possible Rubric for Evaluating Students' Nonfiction Writing

Source: Adapted from "Enhancing Learning Through Formative Assessments and Effective Feedback" (interactive learning module) by J.E. Ormrod, 2015, in *Essentials of Educational Psychology* (4th ed.). Copyright 2015, Pearson. Adapted by permission.

Characteristic	***Proficient***	***In Progress***	***Beginning to Develop***
Correct spelling	Writer correctly spells all words.	Writer correctly spells most words.	Writer incorrectly spells many words.
Correct punctuation & capitalization	Writer uses punctuation marks and uppercase letters where, and only where, appropriate.	Writer occasionally (a) omits punctuation marks, (b) inappropriately uses punctuation marks, or (c) inappropriately uses uppercase/lowercase letters.	Writer makes many punctuation and/or capitalization errors.
Complete sentences	Writer uses complete sentences throughout, except when using an incomplete sentence for a clear stylistic purpose. Writing includes no run-on sentences.	Writer uses a few incomplete sentences that have no obvious stylistic purpose, *or* writer occasionally includes a run-on sentence.	Writer includes many incomplete sentences and/or run-on sentences; writer uses periods rarely or indiscriminately.
Clear focus	Writer clearly states main idea; sentences are all related to this idea and present a coherent message.	Writer only implies main idea; most sentences are related to this idea; a few sentences are unnecessary digressions.	Writer rambles, without a clear main idea; *or* writer frequently and unpredictably goes off topic.
Logical train of thought	Writer carefully leads the reader through his/her own line of thinking about the topic.	Writer shows some logical progression of ideas but occasionally omits a key point essential to the flow of ideas.	Writer presents ideas in no logical sequence.
Convincing statements/ arguments	Writer effectively persuades the reader with evidence or sound reasoning.	Writer includes some evidence or reasoning to support ideas/opinions, but a reader could easily offer counterarguments.	Writer offers ideas/ opinions with little or no justification.

gets scores of 1 on the three writing-mechanics scales and scores of 5 on the three organization-and-logical-flow scales. Both students would have total scores of 18, yet the quality of the students' writing samples would be quite different.

USING TECHNOLOGY

PRACTICAL APPLICATION Computerizing Observations

One good way of enhancing your efficiency in data collection is to record your observations on a laptop, computer tablet, or smartphone as you are making them. For example, when using a checklist, you might create a spreadsheet with a small number of columns—one for each item on the checklist—and a row for every entity you will observe. Then, as you conduct your observations, you can enter an "X" or other symbol into the appropriate cell whenever you see an item in the checklist. Alternatively, you might download free or inexpensive data-collection software for your smartphone or computer tablet; in smartphone lingo, this is called an application, or "app." Examples are OpenDataKit (opendatakit.org) and GIS Cloud Mobile Data Collection (giscloud.com).

For more complex observations, you might create a general *template* document in spreadsheet or word processing software and then electronically "save" a separate version of the document for each person, situation, or other entity you are observing. You can either print out these entity-specific documents for handwritten coding during your observations, or, if time and your keyboarding skills allow, you can fill in each document while on-site in the research setting.

For some types of observations, existing software programs can greatly enhance a researcher's accuracy and efficiency in collecting observational data. An example is CyberTracker (cybertracker.org), with which researchers can quickly record their observations and—using global positioning system (GPS) signals—the specific locations at which they make each observation. For instance, a biologist working in the field might use this software to record specific places at which various members of an endangered animal species or invasive plant species are observed. Furthermore, CyberTracker enables the researcher to custom-design either verbal or graphics-based checklists for specific characteristics of each observation; for instance, a checklist might include photographs of what different flower species look like or drawings of the different leaf shapes that a plant might have.

PRACTICAL APPLICATION Planning and Conducting Interviews in a Quantitative Study

In a quantitative study, interviews tend to be carefully planned in advance, and they are conducted in a similar, standardized way for all participants. Here we offer guidelines for conducting interviews in a quantitative study; some of them are also applicable to the qualitative interviews described in Chapter 9.

GUIDELINES Conducting Interviews in a Quantitative Study

Taking a few simple steps in planning and conducting interviews can greatly enhance the quality of the data obtained, as reflected in the following recommendations.

1. ***Limit questions to those that will directly or indirectly help you answer your research question.*** Whenever you ask people to participate in a research study, you are asking for their *time.* They are more likely to say *yes* to your request if you ask for only a short amount of their time—say, 5 or 10 minutes. If, instead, you want a half hour or longer from each potential participant, you're apt to end up with a sample comprised primarily of people who aren't terribly busy—a potential source of bias that can adversely affect the generalizability of your results.

2. ***As you write the interview questions, consider how you can quantify the responses, and modify the questions accordingly.*** Remember, you are conducting a *quantitative* study. Thus you will, to some extent, be coding people's responses as numbers and, quite possibly, conducting statistical analyses on those numbers. You will be able to assign numerical codes to responses more easily if you identify an appropriate coding scheme ahead of time.

3. ***Restrict each question to a single idea.*** Don't try to get too much information in any single question; in doing so, you may get multiple kinds of data—"mixed messages," so to speak—that are hard to interpret (Gall, Gall, & Borg, 2007).

4. ***Consider asking a few questions that will elicit qualitative information.*** You don't necessarily have to quantify *everything.* People's responses to a few open-ended questions may support or provide additional insights into the numerical data you obtain from more structured questions. By combining quantitative and qualitative data in this manner, you are essentially employing a *mixed-methods design.* Accordingly, we return to the topic of survey research in Chapter 12.

USING TECHNOLOGY

5. ***Consider how you might use a computer to streamline the process.*** Some computer software programs allow you to record interviews directly onto a laptop computer and then transform these conversations into written text (e.g., see Dragon Naturally Speaking; nuance.com/dragon). Alternatively, if interviewees' responses are likely to be short, you might either (a) use a multiple-choice-format checklist to immediately categorize them or (b) directly type them into a spreadsheet or word processing program.

6. ***Pilot-test the questions.*** Despite your best intentions, you may write questions that are ambiguous or misleading or that yield uninterpretable or otherwise useless responses. You can save yourself a great deal of time over the long run if you fine-tune your questions before you begin systematic data collection. You can easily find weak spots in your questions by asking a few volunteers to answer them in a pilot study.

7. ***Courteously introduce yourself to potential participants and explain the general purpose of your study.*** You are more likely to gain potential participants' cooperation if you are friendly, courteous, and respectful and if you explain—up front—what you are hoping to learn in your research. The goal here is to motivate people to *want* to help you out by giving you a little bit of their time.

8. ***Get written permission.*** Recall the discussion of *informed consent* in the section on ethical issues in Chapter 4. All participants in your study (or, in the case of children, their parents or legal guardians) should agree to participate in advance—and in writing.

9. ***Save controversial questions for the latter part of the interview.*** If you will be touching on sensitive topics (e.g., opinions about gun control, attitudes toward people with diverse sexual orientations), put them near the end of the interview, after you have established rapport and gained a person's trust. You might also preface a sensitive topic with a brief statement suggesting that violating certain laws or social norms—although not desirable—is fairly commonplace (Creswell, 2012; Gall et al., 2007). For example, you might say something like this: "Many people admit that they have occasionally driven a car while under the influence of alcohol. Have you ever driven a car when you probably shouldn't have because you've had too much to drink?"

10. ***Seek clarifying information when necessary.*** Be alert for responses that are vague or otherwise difficult to interpret. Simple, nonleading questions—for instance, "Can you tell me more about that?"—may yield the additional information you need (Gall et al., 2007, p. 254).

PRACTICAL APPLICATION Constructing and Administering a Questionnaire

Questionnaires seem so simple, yet in our experience they can be tricky to construct and administer. One false step can lead to uninterpretable data or an abysmally low return rate. We have numerous suggestions that can help you make your use of a questionnaire both fruitful and efficient. We have divided our suggestions into three categories: constructing a questionnaire, using technology to facilitate questionnaire administration and data analysis, and maximizing your return rate.

GUIDELINES Constructing a Questionnaire

Following are 12 guidelines for developing a questionnaire that encourages people to be cooperative and yields responses you can use and interpret. We apologize for the length of the list, but, as we just said, questionnaire construction is a tricky business.

TABLE 6.1 ■ Guide for the Construction of a Questionnaire

Write the question in the space below.	*Why are you asking the question? How does it relate to the research problem?*

what it is intended to measure. Some academic disciplines (e.g., psychology and related fields) insist that a researcher use more formal and objective strategies to determine a questionnaire's validity, especially when the questionnaire is intended to measure complex psychological traits (e.g., personality, motivation, attitudes). We refer you to the section "Determining the Validity of a Measurement Instrument" in Chapter 4 for a refresher on three potentially relevant strategies: creating a table of specifications, taking a multitrait–multimethod approach, and consulting with a panel of experts.

11. ***Scrutinize the almost-final product one more time to make sure it addresses your needs.*** Item by item, a questionnaire should be quality tested again and again for precision, objectivity, relevance, and probability of favorable reception and return. Have you concentrated on the recipient of the questionnaire, putting yourself in the place of someone who is being asked to invest time on your behalf? If you received such a questionnaire from a stranger, would *you* agree to complete it? These questions are important and should be answered impartially.

Above all, you should make sure that *every question is essential for you to address the research problem.* Table 6.1 can help you examine your items with this criterion in mind. Using either paper and pencil or appropriate software (e.g., a spreadsheet or the *table* feature in a word processing program), insert each item in the left-hand column and then, in the right-hand column, explain why you need to include it. If you can't explain how an item relates to your research problem, throw it out!

12. ***Make the questionnaire attractive and professional looking.*** Your final instrument should have clean lines, crystal-clear printing (and certainly no typos!), and perhaps two or more colors. It should ultimately communicate that its author is a careful, well-organized professional who takes his or her work seriously and has high regard for the research participants.

USING TECHNOLOGY

GUIDELINES Using Technology to Facilitate Questionnaire Administration and Data Analysis

Throughout most of the 20th century, questionnaire-based surveys were almost exclusively paper-and-pencil in nature. But with continuing technological advances and people's increasing computer literacy in recent years, many survey researchers are now turning to technology to share some of the burden of data collection and analysis. One possibility is to use a dedicated website both to recruit participants and to gather their responses to survey questions; we address this strategy in a Practical Application feature a bit later in the chapter. Following are several additional suggestions for using technology to make the use of a questionnaire more efficient and cost-effective.

FIGURE 6.8 ■ A Follow-Up Letter

A B C University
Address

Date

Dear [person's name],

We are all very busy these days, and sometimes we have trouble staying on top of our many commitments. Despite our best intentions, we may sometimes overlook something we have said we would do.

Three weeks ago I sent you a questionnaire asking for your input regarding your program at A B C University. To date I have not yet received your completed questionnaire. Perhaps you have simply mislaid it, or perhaps it has been lost in the mail—any one of several reasons might account for its delay in reaching me.

In any event, I am enclosing another copy of the questionnaire, along with another self-addressed, stamped envelope. I am hoping you can find 15 minutes somewhere in your busy schedule to complete and return the questionnaire. I would really appreciate your personal insights and suggestions regarding your experiences in our program.

Thank you once again for your assistance and generosity in helping us enhance our program. And remember that if you have any questions, you can easily reach me at [telephone number] or [e-mail address].

Respectfully yours,

Your Signature

Your Name

USING TECHNOLOGY

PRACTICAL APPLICATION Using the Internet to Collect Data for a Descriptive Study

In recent years, some researchers have collected descriptive data directly on the Internet. For instance, they may put a questionnaire on a website and ask people who visit the site to respond. One site providing links to a wide variety of online research projects is "Psychological Research on the Net," maintained by John Krantz, Professor of Psychology at Hanover College (psych.hanover.edu). As this edition of the book goes to press, the site is hosting research projects on such diverse topics as eating habits, music preferences, religious beliefs, friendships, and parental disciplinary strategies. Dr. Krantz checks to be sure that each project has been approved by the appropriate internal review board and incorporates informed consent procedures. There is no fee for using the site.

Commercial websites for data collection are available as well. Two popular ones are SurveyMonkey (surveymonkey.com) and Zoomerang (zoomerang.com), each of which charges a modest monthly fee. These websites provide templates that make questionnaire design easy and enable a researcher to present a variety of item types (e.g., multiple-choice items, rating scales). They also include features for communicating with a preselected sample of participants (e.g., through e-mail invitations), as well as features through which the researcher can tabulate, statistically analyze, and download the results.

Conducting a survey online has several advantages (Kraut et al., 2004). When the desired sample size is quite large, an online questionnaire is far more cost-effective than a mailed questionnaire. Often a questionnaire can be adapted based on a participant's previous responses; for instance, if a person responds *no* to the question "Do you smoke cigarettes?" the questionnaire software will subsequently skip questions related to smoking habits. Furthermore, some evidence indicates that online surveys yield data comparable to those obtained through face-to-face contact (Gosling, Vazire, Srivastava, & John, 2004).

If you choose to collect data on the Internet, keep in mind that your ethical standards must be just as rigorous as they would be if you were collecting data through face-to-face contacts or the postal service. Participants must be informed about and agree to the general nature of a study, perhaps by means of a website page that serves as an informed consent letter and a virtual "click to accept" button with which participants can indicate consent (Kraut et al., 2004). Also, participants' responses must remain as confidential as they would in any study. The *protection from harm* ethical standard can be especially troublesome in an online study, as it may be virtually impossible to determine that a participant has found a task or question extremely stressful or upsetting and needs some sort of follow-up intervention. Your research advisor and university's internal review board can help you work through ethical issues and develop appropriate precautions for any study that might potentially cause even minor harm or distress to participants.

Sampling, too, must be a source of concern in an online study. SurveyMonkey and Zoomerang enable a researcher to zero in on a predetermined sample of participants—for example, by uploading a list of e-mail addresses to which the participation request will be sent. Other online research projects, such as those on the "Psychological Research on the Net" website mentioned earlier, are open to anyone who wants to participate. But in virtually any online study, the people who participate won't be representative either of a particular group of people or of the overall population of human beings (Gosling et al., 2004; McGraw, Tew, & Williams, 2000). After all, participants will be limited to people who (a) are comfortable with computers, (b) spend a fair amount of time on the Internet, (c) enjoy partaking in research studies, and (d) have been sufficiently enticed by your research topic to do what you ask of them. In cases where a questionnaire can be completed by anyone who has access to the Internet, many responders are apt to be college students who are earning course credit for their participation. In short, *your sample will be biased to some degree.*

Sampling is a concern for any researcher, but it is especially so for the researcher who wants to draw inferences about a large population. In the following section, we look at strategies for selecting an appropriate sample.

CHOOSING A SAMPLE IN A DESCRIPTIVE STUDY

Any researcher who conducts a descriptive study wants to determine *the nature of how things are.* Especially when conducting survey research, the researcher may want to describe one or more characteristics of a fairly large population—perhaps the television viewing habits of 10-year-olds, the teaching philosophies of elementary school teachers, or the attitudes that visitors to Rocky Mountain National Park have about a shuttle bus system. Whether the population is 10-year-olds, elementary school teachers, or national park visitors, we are talking about *very large* groups of people; for example, more than 3 million people visit Rocky Mountain National Park every year.

In such situations, researchers typically do not study the entire population of interest. Instead, they select a subset, or **sample**, of the population. But they can use the results obtained from their sample to make generalizations about the entire population only if *the sample is truly representative of the population.* Here we are talking about a research study's *external validity,* a concept introduced in Chapter 4.

When stating their research problems, many novice researchers forget that they will be studying a sample rather than a population. They announce, for example, that their goal is

> to survey the legal philosophies of the attorneys of the United States and to analyze the relationship of these several philosophical positions with respect to the recent decisions of the Supreme Court of the United States.

If the researcher means what he or she has said, he or she proposes to survey "the attorneys"—all of them! The American Bar Association consists of approximately 400,000 attorneys distributed over more than 3.5 million square miles. Surveying all of them would be a gargantuan undertaking.

A researcher who intends to survey only a subset of a population should say so, perhaps by using such qualifying words as *selected, representative, typical, certain,* or *a random sample of.* For example, the researcher who wants to study the philosophical perspectives of American Bar Association members might begin the problem statement by saying, "The purpose of this research is to survey the legal philosophies of a random sample of attorneys. . . ." Careful researchers say precisely what they mean.

The specific sampling procedure used depends on the purpose of the sampling and a careful consideration of the parameters of the population. But in general, *the sample should be so carefully chosen that, through it, the researcher is able to see characteristics of the total population in the same proportions and relationships that they would be seen if the researcher were, in fact, to examine the total population.*

When you look through the wrong end of a set of binoculars, you see the world in miniature. If the lenses aren't precision-made and accurately ground, you get a distorted view of what you're looking at. In the same way, a sample should, ideally, be a population microcosm. If the sampling procedure isn't carefully planned, any conclusions the researcher draws from the data are likely to be distorted. We discuss this and other possible sources of bias later in the chapter.

Sampling Designs

Different sampling designs may be more or less appropriate in different situations and for different research questions. Here we consider eight approaches to sampling, which fall into two major categories: probability sampling and nonprobability sampling.

Probability Sampling

In **probability sampling**, the sample is chosen from the overall population by *random selection*—that is, it is selected in such a way that each member of the population has an equal chance of being chosen. When such a *random sample* is selected, the researcher can assume that the characteristics of the sample approximate the characteristics of the total population.

An analogy might help. Suppose we have a beaker containing 100 ml of water. Another beaker holds 10 ml of a concentrated acid. We combine the water and acid in proportions of 10:1. After thoroughly mixing the water and acid, we should be able to extract 1 ml from any part of the solution and find that the sample contains 10 parts water for every 1 part acid. In the same way, if we have a population with considerable variability in ethnic background, education level, social standing, wealth, and other factors, and if we have a perfectly selected random sample—a situation usually more theoretical than logistically feasible—we will find in the sample the same characteristics that exist in the larger population, and we will find them in roughly the same proportions.

There are many possible methods of choosing a random sample. For example, we could assign each person in the population a unique number and then use an arbitrary method of picking certain numbers, perhaps by using a roulette wheel (if the entire population consists of 36 or fewer members) or drawing numbers out of a hat. Many computer spreadsheet programs and Internet websites also provide means of picking random numbers (e.g., search for "random number generator").

A popular paper-and-pencil method of selecting a random sample is to use a **table of random numbers**, which you can easily find on the Internet and in many statistics textbooks. Figure 6.9 presents an excerpt from such a table. Typically a table of random numbers includes blocks of digits that can be identified by specific row and column numbers. For instance, the excerpt in Figure 6.9 shows 25 blocks, each of which includes 50 digits arranged in pairs. Each 50-digit block can be identified by both a row number (shown at the very left) and a column number (shown at the very top). To ensure a truly random sample, the researcher identifies a starting point in the table *randomly.*

How might we identify a starting entry number? Pull a dollar bill from your wallet. The one we have just pulled as we write this book has the serial number L45391827A. We choose the first 2 digits of the serial number, which makes the entry number 45. But which is the row the center of a particular city. As people pass, you interview them. The fact that people in the two categories may come in clusters of two, three, or four is no problem. All you need are the opinions of 20 people from each category. This type of sampling regulates only the size of each category within the sample; in every other respect, the selection of the sample is nonrandom and, in most cases, convenient.

Purposive Sampling In purposive sampling, people or other units are chosen, as the name implies, for a particular *purpose.* For instance, we might choose people who we have decided are "typical" of a group or those who represent diverse perspectives on an issue.

Pollsters who forecast elections frequently use purposive sampling: They may choose a combination of voting districts that, in past elections, has been quite helpful in predicting the final outcomes.

Purposive sampling may be very appropriate for certain research problems. However, researchers should always provide a rationale explaining why they selected their particular sample of participants.

Sampling in Surveys of Very Large Populations

Nowhere is sampling more critical than in surveys of large populations. Sometimes a researcher reports that *x*% of people believe such-and-such, that *y*% do so-and-so, or that *z*% are in favor of a particular political candidate. *Such percentages are meaningless unless the sample is representative of the population about which inferences are to be drawn.*

But now imagine that a researcher wants to conduct a survey of the country's *entire adult population.* How can the researcher possibly hope to get a random, representative sample of such a large group of people? The Survey Research Center of the University of Michigan's Institute for Social Research has successfully used a *multistage sampling of areas,* described in its now-classic *Interviewer's Manual* (1976):

1. ***Primary area selection.*** The country is divided into small "primary areas," each consisting of a specific county, a small group of counties, or a large metropolitan area. A predetermined number of these areas are randomly selected.

2. ***Sample location selection.*** Each of the selected primary areas is divided into smaller sections ("sample locations"), such as specific towns. A small number of these locations is randomly selected.

3. ***Chunk selection.*** The sample locations are divided into even smaller "chunks" that have identifiable boundaries such as roads, streams, or the edges of a city block. Most chunks have 16 to 50 dwellings, although the number may be larger in large cities. Once again, a random sample is selected.

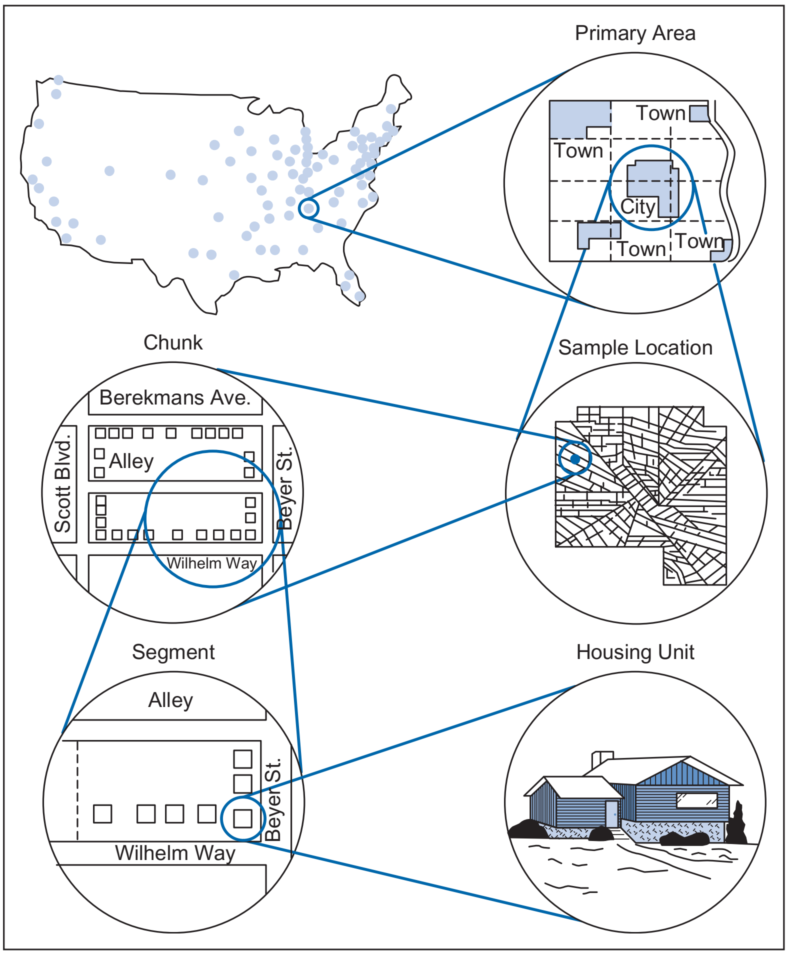
4. ***Segment selection.*** Chunks are subdivided into areas containing a relatively small number of dwellings, and some of these "segments" are, again, chosen randomly.

5. ***Housing unit selection.*** Approximately four dwellings are selected (randomly, of course) from each segment, and the residents of those dwellings are asked to participate in the survey. If a doorbell is unanswered, the researcher returns at a later date and tries again.

As you may have deduced, the approach just described is a multistage version of cluster sampling (see Figure 6.14). At each stage of the game, units are selected randomly. "Randomly" does *not* mean haphazardly or capriciously. Instead, a mathematical procedure is employed to ensure that selection is entirely random and the result of blind chance. This process should yield a sample that is, in all important respects, representative of the country's population.

FIGURE 6.14 Multistage Sampling

Source: From the *Interviewer's Manual* (Rev. ed., p. 36) by the Survey Research Center, Institute for Social Research, 1976, Ann Arbor: University of Michigan. Reprinted with permission.



PRACTICAL APPLICATION Identifying a Sufficient Sample Size

A basic rule in sampling is: *The larger the sample, the better.* But such a generalized rule isn't very helpful to a researcher who must make a practical decision about a specific research situation. Gay, Mills, and Airasian (2012, p. 139) have offered the following guidelines for selecting a sample size, which we'll refer to by the symbol *N*:

- For smaller populations, say, $N = 100$ or fewer, there is little point in sampling; survey the entire population.
- If the population size is around 500 (give or take 100), 50% should be sampled.
- If the population size is around 1,500, 20% should be sampled.
- Beyond a certain point (about $N = 5{,}000$), the population size is almost irrelevant and a sample size of 400 will be adequate.

Generally speaking, then, the larger the population, the smaller the percentage—but not the smaller the number!—one needs to get a representative sample.

To some extent, the size of an adequate sample depends on how homogeneous or heterogeneous the population is—how alike or different its members are with respect to the characteristics of research interest. If the population is markedly heterogeneous, a larger sample is necessary than if the population is fairly homogeneous. Important, too, is the degree of precision with which the researcher wants to draw conclusions or make predictions about the population under study.

Statisticians have developed formulas for determining the desired sample size for a given population. Such formulas are beyond the scope of this book, but you can find them in many introductory statistics books and on many Internet websites (e.g., search "calculating sample size").

PRACTICAL APPLICATION Analyzing the Population in a Descriptive Study

Select a particular population and conduct an analysis of its structure and characteristics. Analyze the population you have chosen by completing the following checklist.

CHECKLIST

Analyzing Characteristics of the Population Being Studied

_____ 1. On the following line, identify the particular population you have chosen:

____

_____ 2. Now answer the following questions with respect to the *structure of the population:*

	YES	NO
a. Is the population a relatively homogeneous group of individuals or other units?	_____	_____
b. Could the population be considered to consist generally of equal "layers," each of which is fairly homogeneous in composition?	_____	_____
c. Could the population be considered to be composed of separate homogeneous layers differing in size and number of units comprising them?	_____	_____
d. Could the population be envisioned as isolated islands or clusters of individual units, with the clusters being similar to one another in composition?	_____	_____

_____ 3. Through what means would you extract a representative sample from the total population? Describe your procedure on the following lines:

____

____

____

_____ 4. Refer to Table 6.2. Is your sampling procedure appropriate for the characteristics of the population? _____ Yes _____ No

_____ 5. Have you guaranteed that your sample will be chosen by chance and yet will be representative of the population? _____ Yes _____ No

_____ 6. If the preceding answer is yes, explain how this will be done.

____

____

____

______ 7. Indicate what means will be employed to obtain the information you need from the sample.

____

____

______ 8. What are the weaknesses inherent in this method of obtaining the data?

____

____

____

COMMON SOURCES OF BIAS IN DESCRIPTIVE STUDIES

In this and preceding chapters, we have occasionally mentioned that a particular research strategy might in some way *bias* the results. In general, **bias** in a research study is any influence, condition, or set of conditions that singly or in combination distort the data obtained or conclusions drawn. Bias can creep into a research project in a variety of subtle ways. For example, when conducting an interview, a researcher's tone of voice in asking questions might predispose a participant to respond in one way rather than in another, or the researcher's personality might influence a participant's willingness to reveal embarrassing facts.

Most sources of bias in descriptive research fall into one of four categories, each of which we examine now.

Sampling Bias

A key source of bias in many descriptive studies is **sampling bias**—any factor that yields a nonrepresentative sample of the population being studied. For example, imagine that a researcher wants to conduct a survey of a certain city's population and decides to use the city telephone book as a source for selecting a random sample. She opens to a page at random, closes her eyes, puts her pencil down on the page, and selects the name that comes closest to the pencil point. "You can't get more random than this," she thinks. But the demon of bias is there. Her possible selections are limited to people who are listed in the phone book. People with very low income levels won't be adequately represented because some of them can't afford telephone service. Nor will wealthy individuals be proportionally represented because many of them have unlisted numbers. And, of course, people who use only cell phones—people who, on average, are fairly young—aren't included in the phone book. Hence, the sample will consist of disproportionately large percentages of people at middle-income levels and in older age-groups (e.g., Keeter, Dimock, Christian, & Kennedy, 2008). Likewise, as noted in earlier sections of the chapter, studies involving online interviews or Internet-based questionnaires are apt to be biased—this time in favor of computer-literate individuals with easy access to the Internet.

Studies involving mailed questionnaires frequently fall victim to bias as well, often without the researcher's awareness. For example, suppose that a questionnaire is sent to 100 citizens, asking, "Have you ever been audited by the Internal Revenue Service (IRS) to justify your income tax return?" Of the 70 questionnaires returned, 35 are from people who say that they have been audited, whereas 35 are from people who respond that they have never been audited. The researcher might therefore conclude that 50% of American citizens are likely to be audited by the IRS at one time or another.

The researcher's generalization isn't necessarily accurate. We need to consider how the nonrespondents—30% of the original sample—might be different from those who responded to the questionnaire. Many people consider an IRS audit to be a reflection of their integrity. Perhaps for this reason, some individuals in the researcher's sample may not have wanted to admit that they had been audited and so tossed the questionnaire into the wastebasket. If previously audited people were less likely to return the questionnaire than nonaudited people, the sample was biased, and thus the results didn't accurately represent the facts. Perhaps, instead of a 50-50 split, an estimate of 60% (people audited) versus 40% (people not audited) is more accurate. The data the researcher has obtained don't enable the researcher to make such an estimate, however.

The examples just presented illustrate two different ways in which bias can creep into the research sample. In the cases of telephone and Internet-based data collection, *sample selection* itself was biased because not everyone in the population had an equal chance of being selected. For instance, people not listed in the phone book had *zero* chance of being selected. Here we see the primary disadvantage of nonprobability sampling, and especially of convenience sampling: People who happen to be readily available for a research project—those who are in the right place at the right time—are almost certainly *not* a random sample of the overall population.

In the example concerning IRS audits, *response rate*—and, in particular, potential differences between respondents and nonrespondents—was the source of bias. In that situation, the researcher's return rate of 70% was quite high. More often, the return rate in a questionnaire study is 50% or less, and the more nonrespondents there are, the greater the likelihood of bias. Likewise, in telephone surveys, a researcher won't necessarily reach certain people even with 10 or more attempts, and those who *are* eventually reached won't all agree to an interview (Witt & Best, 2008).

Nonrespondents to *mailed questionnaires* might be different from respondents in one or more ways (Rogelberg & Luong, 1998). They may have illnesses, disabilities, or language barriers that prevent them from responding. And on average, they have lower educational levels. In contrast, people who are hard to reach *by telephone* are apt to be young working adults who are *more* educated than the average individual (Witt & Best, 2008).

Even when potential participants' ages, health, educational levels, language skills, and computer literacy are similar, they can differ widely in their *motivation* to participate in a study: Some might have other priorities, and some might worry that a researcher has sinister intentions. Participants in longitudinal studies may eventually grow weary of being "bothered" time after time. Also, a nonrandom subset of them might die before the study is completed!

Look once again at the five steps in the University of Michigan's Survey Research Center procedure for obtaining a sample in a national survey. Notice the last sentence in the fifth step: "If a doorbell is unanswered, the researcher returns at a later date and tries again." The researcher does *not* substitute one housing unit for another; doing so would introduce bias into the sampling design. The center's *Interviewer's Manual* describes such bias well:

> The house on the muddy back road, the apartment at the top of a long flight of stairs, the house with the growling dog outside must each have an opportunity to be included in the sample. People who live on back roads can be very different from people who live on well paved streets, and people who stay at home are not the same as those who tend to be away from home. If you make substitutions, such important groups as young men, people with small families, employed women, farmers who regularly trade in town, and so on, may not have proportionate representation in the sample. (Survey Research Center, 1976, p. 37)

Instrumentation Bias

By **instrumentation bias**, we mean the ways in which particular measurement instruments slant the obtained results in one direction or another. For instance, in our earlier discussion of questionnaires, we mentioned that a researcher must choose certain questions—and by default must omit *other* questions. The same is true of structured interviews: By virtue of the questions asked, participants are encouraged to reflect on and talk about some topics rather than other ones. The outcome is that some variables are included in a study, and other potentially important variables are overlooked.

As an example, imagine that an educational researcher is interested in discovering the kinds of goals that students hope to accomplish when they're at school. Many motivation researchers have speculated that students might be concerned about either (a) truly mastering classroom subject matter, on the one hand, or (b) getting good grades by any expedient means, on the other. Accordingly, they have designed and administered rating-scale questionnaires with such items as "I work hard to understand new ideas" (reflecting a desire to master a topic) and "I occasionally copy someone else's homework if I don't have time to do it myself" (reflecting a desire to get good grades). But in one study (Dowson & McInerney, 2001), researchers instead asked middle students what things were most important for them to accomplish at school. Many participants focused not on a desire to do well academically but instead on *social* goals, such as being with and helping classmates and avoiding behaviors that might adversely affect their popularity.

Response Bias

Whenever we gather data through interviews or questionnaires, we are relying on *self-report* data: People are telling us what they believe to be true or, perhaps, what they think we want to hear. To the extent that people describe their thoughts, beliefs, and experiences inaccurately, **response bias** is at work. For example, people's descriptions of their attitudes, opinions, and motives are often constructed on the spot—sometimes they haven't really thought about a certain issue until a researcher poses a question about it—and thus may be colored by recent events, the current context, or flawed self-perceptions (McCaslin, Vega, Anderson, Calderon, & Labistre, 2011; Schwarz, 1999). Furthermore, some participants may intentionally or unintentionally misrepresent the facts in order to give a favorable impression—a source of bias known as a *social desirability effect* (e.g., Uziel, 2010). For example, if we were to ask parents the question, "Have you ever abused your children?" the percentage of parents who told us *yes* would be close to zero, and so we would almost certainly underestimate the prevalence of child abuse in our society. And when we ask people about *past* events, behaviors, and perspectives, interviewees must rely on their memories, and human memory is rarely as accurate as a video recorder might be. People are apt to recall what *might* or *should* have happened (based on their attitudes or beliefs) rather than what actually *did* happen (e.g., Schwarz, 1999; Wheelan, 2013).

Researcher Bias

Finally, we must not overlook the potential effects of a researcher's expectations, values, and general belief systems, which can predispose the researcher to study certain variables and not other variables, as well as to draw certain conclusions and not other conclusions. For example, recall the discussion of philosophical assumptions in Chapter 1: Researchers with a *positivist* outlook are more likely to look for cause-and-effect relationships—sometimes even from correlational studies that don't warrant conclusions about cause and effect!—than postpositivists or constructivists.

Ultimately, we must remember that *no human being can be completely objective.* Assigning numbers to observations helps a researcher quantify data but it does not necessarily make the researcher any more objective in collecting or interpreting those data.

PRACTICAL APPLICATION Acknowledging the Probable Presence of Bias in Descriptive Research

When conducting research, it's almost impossible to avoid biases of one sort or another—biases that can potentially influence the data and thus also influence the conclusions drawn. Good researchers demonstrate their integrity by admitting, without reservation, that certain biases may well have influenced their findings. For example, in survey research, you should *always* report the percentages of people who have and have not consented to participate, such as those who have agreed and refused to be interviewed or those who have and have not returned questionnaires. Furthermore, you should be candid about possible sources of bias that result from differences between participants and nonparticipants. Here we offer guidelines for identifying possible sampling biases in questionnaire research. We then provide a checklist that can help you pin down various biases that can potentially contaminate descriptive studies of all sorts.

_______ 2. Do you have any preconceived notions about cause-and-effect relationships within the phenomenon you are studying? If so, what precautions might you take to ensure that you do *not* infer causal relationships from cross-variable correlations you might find?

____

____

_______ 3. How do you plan to identify a sample for your study? What characteristics of that sample might limit your ability to generalize your findings to a larger population?

____

____

_______ 4. On what specific qualities and characteristics will you be focusing? What potentially relevant qualities and characteristics will you *not* be looking at? To what degree might omitted variables be as important or more important in helping to understand the phenomenon you are studying?

____

____

_______ 5. Might participants' responses be poor indicators of certain characteristics, attitudes, or opinions? For example:

- Might they say or do things in order to create a favorable impression?

  _______ Yes _______ No

- Might you be asking them questions about topics they haven't really thought about before?

  _______ Yes _______ No

- Will some questions require them to rely on their memories of past events?

  _______ Yes _______ No

If any of your answers are *yes,* how might such sources of bias influence your findings?

____

____

INTERPRETING DATA IN DESCRIPTIVE RESEARCH

In our discussion of descriptive research methods in this chapter, we have focused largely on strategies for acquiring data. But at this juncture, we remind you of two basic principles of research:

1. The purpose of research is to seek the answer to a problem in light of data that relate to the problem.
2. Although collecting data for study and organizing it for inspection require care and precision, extracting meaning from the data—the interpretation of the data—is all-important.

A descriptive study is often a very "busy" research method: The researcher must decide on a population; choose a technique for sampling it; develop a valid means of collecting the desired information; minimize the potential for bias in the study; and then actually collect, record, organize, and analyze all the necessary data. The activities connected with descriptive research can be complex, time-consuming, and occasionally distracting. Therein lies an element of danger. With all this action going on, it wouldn't be surprising if the researcher lost sight of the problem and subproblems. But the problem and its subproblems are precisely the reason for the entire endeavor.

Activity for activity's sake is seductive. Amassing great quantities of data can provide a sense of well-being, and a researcher might lose sight of the ultimate demands that the problem itself makes on those data. Presenting the data in displays and summaries—graphs, charts, tables—does nothing more than demonstrate the researcher's acquisitive skills and consummate ability to present the same data in various ways.

All research activity is subordinate to the research problem itself. Sooner or later, the entire effort must result in an interpretation of the data and a setting forth of conclusions, drawn from the data, to resolve the problem under investigation. Descriptive research ultimately aims to solve problems through the *interpretation* of the data that have been gathered.

SOME FINAL SUGGESTIONS

As we approach the end of the chapter, it is important to reflect on several issues related to descriptive research. Consider each of the following questions within the context of the research project you have in mind:

- Why is a description of this population and/or phenomenon valuable?
- What specific data will I need to solve my research problem and its subproblems?
- What procedures should I follow to obtain the necessary information? How can I best implement those procedures?
- How do I get a sample that will be reflective of the entire population about which I am concerned?
- How can I collect my data in a way that minimizes misrepresentations and misunderstandings?
- How can I control for possible bias in the collection and description of the data?
- What do I do with the data once I have collected them? How do I organize and prepare them for analysis?
- Above all, in what ways might I reasonably interpret the data? What conclusions might I reach from my investigation?

A SAMPLE DISSERTATION

We conclude the chapter by illustrating how questionnaires might be used in a correlational study to address the topic of violence in intimate relationships (e.g., husband and wife, boyfriend and girlfriend) in American society. The excerpts we present are from Luis Ramirez's doctoral dissertation in sociology completed at the University of New Hampshire (Ramirez, 2001).

Ramirez hypothesized that violence between intimate partners—in particular, assault by one partner on the other—is, in part, a function of ethnicity, acculturation (e.g., adoption of mainstream American behaviors and values), criminal history, and social integration (e.g., feelings of connectedness with family and friends). He further hypothesized that as a result of such factors, differences in intimate partner violence might be observed in Mexican Americans and non-Mexican Americans.

Ramirez begins Chapter 1 by discussing the prevalence of violence (especially assault) in intimate relationships. We pick up Chapter 1 at the point where he identifies his research questions and hypotheses. We then move into Chapter 2, where he describes his methodology. As has been true for earlier proposal and dissertation samples, the research report appears on the left-hand side, and our commentary appears on the right.

FOR FURTHER READING

Alreck, P. L., & Settle, R. B. (2003). *The survey research handbook* (3rd ed.). New York: McGraw-Hill.

Berdie, D. R., Anderson, J. F., & Niebuhr, M. A. (1986). *Questionnaires: Design and use* (2nd ed.). Metuchen, NJ: Scarecrow Press.

Bourque, L. B., & Fielder, E. P. (Eds.). (2002a). *How to conduct self-administered and mail surveys* (2nd ed.). Thousand Oaks, CA: Sage.

Bourque, L. B., & Fielder, E. P. (2002b). *How to conduct telephone surveys* (2nd ed.). Thousand Oaks, CA: Sage.

Daniel, J. (2011). *Sampling essentials: Practical guidelines for making sampling choices.* Thousand Oaks, CA: Sage.

Delandshere, G., & Petrosky, A. R. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher, 27,* 14–24.

Fink, A. (2009). *How to conduct surveys: A step-by-step guide* (5th ed.). Thousand Oaks, CA: Sage.

Fowler, F. J., Jr. (2014). *Survey research methods* (5th ed.). Thousand Oaks, CA: Sage.

Friedman, H. H., & Amoo, T. (1999). Rating the rating scales. *Journal of Marketing Management, 9*(3), 114–123.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59,* 93–104.

Gubrium, J. F., Holstein, J. A., Marvasti, A. B., & McKinney, K. D. (Eds.). (2012). *The SAGE handbook of interview research: The complexity of the craft* (2nd ed.). Thousand Oaks, CA: Sage.

Gwartney, P. A. (2007). *The telephone interviewer's handbook: How to conduct standardized conversations.* San Francisco, CA: Jossey-Bass.

Henry, G. T. (1990). *Practical sampling.* Thousand Oaks, CA: Sage.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. *American Psychologist, 59,* 105–117.

Laursen, B., Little, T. D., & Card, N. A. (Eds.). (2013). *Handbook of developmental research methods.* New York: Guilford Press.

Litwin, M. S. (1995). *How to measure survey reliability and validity.* Thousand Oaks, CA: Sage.

Magnusson, D., Bergman, L. R., Rudinger, G., & Torestad, B. (Eds.). (2010). *Problems and methods in longitudinal research: Stability and change.* New York: Cambridge University Press.

Marsden, P. V., & Wright, J. D. (Eds.). (2010). *Handbook of survey research* (2nd ed.). Bingley, United Kingdom: Emerald Group.

Menard, S. (Ed.). (2007). *Handbook of longitudinal research: Design, measurement, and analysis.* Burlington, MA: Academic Press.

Neuman, W. L. (2011). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Boston: Allyn & Bacon. (See Chapter 10, "Survey Research.")

Oppenheim, A. N. (1992). *Questionnaire design, interviewing, and attitude measurement.* New York: St. Martin's Press.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38,* 109–119.

Rogelberg, S. G., & Luong, A. (1998). Nonresponse to mailed surveys: A review and guide. *Current Directions in Psychological Science, 7,* 60–65.

Schuman, H., & Presser, S. (1996). *Questions and answers in attitude surveys: Experiments on question form, wording, and context.* Thousand Oaks, CA: Sage.

Schwarz, P. N., & Sudman, S. (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Francisco: Jossey-Bass.

Sue, V. M., & Ritter, L. A. (2012). *Conducting online surveys* (2nd ed.). Thousand Oaks, CA: Sage.

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples.* New York: Springer.

Chapter 7

Experimental, Quasi-Experimental, and Ex Post Facto Designs

Progress is relative: We measure it by noting the amount of change between what was and what is. And we attempt to account for the change by identifying the dynamics that have caused it. Ideally, we must manipulate one possible causal factor while controlling all other possible causal factors; only in this way can we determine whether the manipulated factor has a direct influence on the phenomenon we are studying. To the extent that many potentially causal factors all vary at once in an entangled, confounded manner, we learn little or nothing about what causes what.

Learning Outcomes

7.1 Identify examples of independent and dependent variables, and describe several strategies for controlling for confounding variables in experimental studies.

7.2 Recognize examples of pre-experimental, experimental, quasi-experimental, ex post facto, and factorial designs, as well as of meta-analyses.

7.3 Describe potential biases in these designs and how they might affect a study's internal or external validity.

In the descriptive designs described in the preceding chapter, we make no systematic attempt to determine the underlying causes of the phenomena being studied. But sometimes we *do* want to know what leads to what; in other words, we want to identify *cause-and-effect relationships.*

A researcher can most convincingly identify cause-and-effect relationships by using an **experimental design**. In such a design, the researcher considers many possible factors that might cause or influence a particular condition or phenomenon. The researcher then attempts to control for all influential factors *except* those whose possible effects are the focus of investigation.

An example can help clarify the point. Imagine that we have two groups of people. We take steps to make sure that, on average, the two groups are so similar that we can, for all intents and purposes, call them equivalent. We give members of both groups a pretest to measure a particular characteristic in which we are interested—for instance, this might be blood pressure, academic achievement, or purchasing habits. Then we expose only one of the groups to a **treatment** or intervention of some sort—perhaps a new pharmaceutical drug, an instructional method, or an advertising campaign—that we think may have an effect on the characteristic we are studying. Afterward, we give members of both groups a posttest to measure the characteristic once again. If the characteristic changes for the group that received the intervention but does *not* change for the other group, and if everything about the two groups has been the same *except for the intervention,* we can reasonably conclude that the treatment or intervention brought about the change we observed. Because we have systematically *manipulated* the situation, we have used an experimental design.

Some of the research designs we describe in this chapter are true experimental designs; as such, they allow us to identify cause-and-effect relationships. Other designs in this chapter eliminate some—but not all—alternative explanations of an observed change. Yet all of the designs in the chapter have one thing in common: clearly identifiable independent and dependent variables.

We have previously introduced you to independent and dependent variables in Chapter 2, but because these concepts guide so much of our discussion in this chapter, a brief refresher might be in order. An **independent variable** is one that the researcher studies as having a possible effect on one or more other variables. In many of the designs described in this chapter, the researcher directly manipulates and controls at least one independent variable. In contrast, a **dependent variable** is a variable that is potentially influenced by an independent variable; that is, its value *depends* to some degree on the value of the independent variable. In other words, the hypothesized relationship is this:

$$\text{Independent variable} \rightarrow \text{Dependent variable}$$

As an example, let's look at a dissertation in educational psychology written by Nancy Thrailkill (1996), who wanted to study the effects of three different kinds of lecture material on people's ability to remember information contained in the lecture. Working with undergraduate students, she presented different parts of a lecture on an obscure American Civil War battle in one of three ways: (a) she described certain historical figures and events in such a manner that they were easy to imagine and visualize (*imagery* condition), (b) she included attention-grabbing phrases in the lecture (*attention* condition), or (c) she did neither of these things (*control* condition). In the following examples from Thrailkill's dissertation, the underscored phrases illustrate the modifications made for each of the three conditions; other variations in wording made the three lectures equivalent in length:

> *Imagery:* Lincoln also created the Army of Virginia, incorporating several forces which had been under different commanders. Lincoln set <u>the dimpled, baby-faced young blond</u> Major General John Pope in charge of this new combined force. Being put under his command was objectionable to some of the former commanders. . . .
>
> *Attention:* Lincoln also created the Army of Virginia, incorporating several forces which had been under different commanders. <u>LISTEN TO ME NOW.</u> Lincoln set the less experienced Major General John Pope in charge of this new combined force. Being put under the command of Pope was objectionable to some of the former commanders. . . .
>
> *Control:* Lincoln also created the Army of Virginia, incorporating several forces which had been under different commanders. Lincoln set the less experienced <u>junior officer</u> Major General John Pope in charge of this new combined force. Being put under the command of Pope was objectionable to some of the former commanders. (Thrailkill, 1996, p. 62, some underscoring added)

After presenting different parts of the lecture under the three different conditions, Thrailkill measured the students' recall for the lecture in two ways. She first gave students blank sheets of paper and asked them to write down as much of the lecture as they could remember (a *free recall* task). When they had completed that task, she gave them a multiple-choice test that assessed their memory for specific facts within the lecture. In this study, the independent variable was the nature of the lecture material: easily visualized, attention-getting, or neutral. There were two dependent variables, both of which reflected students' ability to recall facts within the lecture: students' performance on the free recall task and their scores on the multiple-choice test. Thrailkill's hypothesis was confirmed: The students' ability to recall lecture content *depended,* to some extent, on the way in which the content was presented.

THE IMPORTANCE OF CONTROL

A particular concern in any experimental study is its **internal validity**, the extent to which its design and the data it yields allow the researcher to draw legitimate conclusions about cause-and-effect and other relationships (see Chapter 4). In experimental designs, internal validity is essential. Without it, a researcher cannot draw firm conclusions about cause and effect—and that is, after all, the whole point of conducting an experimental study.

As an example, suppose we have just learned about a new method of teaching science in elementary school. We want to conduct an experiment to investigate the method's effect on students' science achievement test scores. We find two fifth-grade teachers who are willing to participate in the study. One teacher agrees to use the new method in the coming school year; in fact, she's quite eager to try it. The other teacher wants to continue using the same approach he has always used. Both teachers agree that at the end of the school year we can give their students a science achievement test.

Are the two classes the same in every respect *except for the experimental intervention?* If the students taught with the new method obtain higher achievement test scores at the end of the year, will we know that the method was the *cause* of the higher scores? The answer to both questions is a resounding *no!* The teachers are different: One is female and the other male, and they almost certainly have different personalities, educational backgrounds, and teaching styles. In addition, the two groups of students may be different; perhaps the students instructed by the new method are, on average, more intelligent or motivated than the other, or perhaps they live in a more affluent school district. Other, more subtle differences may be at work as well, including the interpersonal dynamics in the two classes, and the light, temperature, and noise levels within each classroom. Any of these factors—and perhaps others we haven't thought of—might be reasons for any group differences in achievement test scores we obtain.

Whenever we compare two or more groups that are or might be different in ways *in addition to* the particular treatment or intervention we are studying, we have **confounding variables** in our study. The presence of such variables makes it extremely difficult to draw conclusions about cause-and-effect relationships, because we can't pin down exactly *what* is the cause of any pattern in the data observed after the intervention. In other words, confounding variables threaten a study's internal validity. In a now-classic book chapter, Campbell and Stanley (1963) identified several potential threats to the internal validity of an experimental study; we describe them in Figure 7.1.

Controlling for Confounding Variables

To maximize internal validity when a researcher wants to identify cause-and-effect relationships, the researcher needs to control confounding variables in order to rule them out as explanations for any effects observed. Researchers use a variety of strategies to control for confounding variables. Following are several common ones.

1. ***Keep some things constant.*** When a factor is the *same* for everyone, it cannot possibly account for any differences observed. Oftentimes researchers ensure that different treatments are imposed in the same or very similar environments. They may also seek research participants who share a certain characteristic, such as age, gender, grade level, or socioeconomic status. Keep in mind, however, that restricting the nature of one's sample may lower the *external validity,* or generalizability, of any findings obtained (see the discussion of this concept in Chapter 4).

2. ***Include a control group.*** In Chapter 4 we described a study in which an industrial psychologist begins playing classical music as employees in a typing pool go about their daily task of typing documents. At the end of the month, the psychologist finds that the typists' productivity is 30% higher than it was during the preceding month. The increase in productivity may or may not be due to the classical music. There are too many possible confounding variables—personnel changes, nature of the documents being typed, numbers of people out sick or on vacation during the 2-month period, even just the knowledge that an experiment is being conducted—that may also account for the typing pool's increased productivity.

To better control for such extraneous variables, researchers frequently include a **control group**, a group that receives either no intervention or a "neutral" intervention that should have little or no effect on the dependent variable. The researchers then compare the performance of this group to an **experimental group**—also known as a **treatment group**—that participates in an intervention.

6. ***Statistically control for confounding variables.*** Sometimes researchers can control for known confounding variables, at least in part, through statistical techniques. Such techniques as *partial correlation, analysis of covariance* (ANCOVA), and *structural equation modeling* (SEM) are suitable for this purpose. We briefly describe each of these in Chapter 8. Should you choose to use one of them in your own research, we urge you to consult one or more statistics books for guidance about their use and appropriateness for various research situations.

Keep in mind, however, that statistically controlling confounding variables is no substitute for controlling for them in one's research design if at all possible. *A carefully controlled experimental design is the only approach that allows you to draw firm conclusions about cause and effect.*

OVERVIEW OF EXPERIMENTAL, QUASI-EXPERIMENTAL, AND EX POST FACTO DESIGNS

In true experimental research, the researcher manipulates the independent variable and examines its effects on another, dependent variable. A variety of research designs have emerged that differ in the extent to which the researcher manipulates the independent variable and controls for confounding variables—in other words, the designs differ in the degree to which they have *internal validity.* In the upcoming sections, we present a number of possible designs, which we have divided into five general categories: *pre-experimental designs, true experimental designs, quasi-experimental designs, ex post facto designs,* and *factorial designs.* Altogether we describe 16 different designs that illustrate various ways—some more effective than others—of attempting to identify cause-and-effect relationships. Some of our discussion is based on designs identified by Campbell and Stanley (1963).[1]

We illustrate the designs using tables that have this general format:

Group	Time →		
Group 1			
Group 2			

Each group in a design is shown in a separate row, and the things that happen to the group over time are shown in separate cells within the row. The cells have one of four notations:

Tx: Indicates that a *treatment* (reflecting the independent variable) is presented.

Obs: Indicates that an *observation* (reflecting the dependent variable) is made.

—: Indicates that nothing occurs during a particular time period.

Exp: Indicates a previous *experience* (an independent variable) that some participants have had and others have not; the experience has *not* been one that the researcher could control.

The nature of these tables will become more apparent as we proceed.

As you read about the 16 designs, keep in mind that they are hardly an exhaustive list; researchers can modify or combine them in various ways. For example, although we will be limiting ourselves to studies with only one or two groups (perhaps one treatment group and one control group), it's entirely possible to have two or more treatment groups (each of which is exposed to a different variation of the independent variable) and, in some cases, two control groups (perhaps one getting a placebo and another getting no intervention at all). More generally, the designs we describe here should simply provide starting points that get you thinking about how you might best tackle your own research problem.

[1]In particular, Designs 1 to 6 and Designs 8 to 11 are based on those that Campbell and Stanley described. However, when describing Design 11, we use the contemporary term *reversal time-series design* rather than Campbell and Stanley's original term *equivalent time-samples design.*

PRE-EXPERIMENTAL DESIGNS

In **pre-experimental designs**, it isn't possible to show cause-and-effect relationships, because either (a) the independent "variable" doesn't vary or (b) experimental and control groups are not comprised of equivalent or randomly selected individuals. Such designs are helpful only for forming tentative hypotheses that should be followed up with more controlled studies.

Design 1: One-Shot Experimental Case Study

The one-shot experimental case study is probably the most primitive type of experiment that might conceivably be termed "research." An experimental treatment (Tx) is introduced, and then a measurement (Obs)—a posttest of some sort—is administered to determine the effects of the treatment. This design is shown in the following table:

Group	Time →	
Group 1	Tx	Obs

The design has low internal validity because it's impossible to determine whether participants' performance on the posttest is the result of the experimental treatment per se. Many other variables may have influenced participants' performance, such as physiological maturation or experiences elsewhere in the participants' general environment. Perhaps the characteristic or behavior observed after the treatment existed *before* the treatment as well. The reality is that with a single measurement or observation, we have no way of knowing whether the situation has changed or not, let alone whether it has changed as a result of the intervention.

One-shot experimental case studies may be at the root of many common misconceptions. For example, imagine that we see a child sitting on the ground on a damp, rainy day. The next day the child has a sore throat and a cold. We conclude that sitting on the damp earth caused the child to catch cold. Thus, the design of our "research" thinking is something like this:

Exposure to cold, damp ground (Tx) → Child has a cold (Obs)

Such "research" may also "support" such superstitious folk beliefs as these: If you walk under a ladder, you will have bad luck; Friday the 13th is a day of catastrophes; a horseshoe above the front door brings good fortune to one's home. Someone observed an event, then observed a subsequent event, and linked the two together as cause and effect.

Be careful not to confuse the one-shot experimental case study method with the qualitative case study design described in Chapter 9. Case study research involves extensive engagement in a research setting—a far cry from basing conclusions on a single observation.

Although the one-shot experimental case study is simple to carry out, its results are, for all intents and purposes, meaningless. At the very least, researchers should use the design described next.

Design 2: One-Group Pretest–Posttest Design

In a one-group pretest–posttest design, a single group (a) undergoes a pre-experimental observation or evaluation, then (b) is administered the experimental treatment, and finally (c) is observed or evaluated again after the treatment. This design can be represented as follows:

Group	Time →		
Group 1	Obs	Tx	Obs

Suppose an elementary school teacher wants to know if simultaneously reading a story and listening to it on audiotape will improve the reading skills of students in his class. He gives his students a standardized reading test, then has them simultaneously read and listen to simple stories every day for 8 weeks, and then administers an alternate form of the same standardized reading test. If the students' test scores improve over the 8-week period, the teacher might conclude—perhaps accurately, but perhaps not—that the simultaneous-reading-and-listening intervention was the cause of the improvement.

Now suppose an agronomist crossbreeds two strains of corn. She finds that the resulting hybrid strain is more disease-resistant and has a better yield than either of the two parent types. She concludes that the crossbreeding process has made the difference. Once again we have an Obs–Tx–Obs design: The agronomist measures the disease level of the parent strains (Obs), then develops a hybrid of the two strains (Tx), and then measures the disease level of the next generation (Obs).

In a one-group pretest–posttest design, we at least know that a change has taken place. However, we haven't ruled out other possible explanations for the change. In the case of the elementary school teacher's study, improvement in reading scores may have been due to other activities within the classroom curriculum, to more practice taking the reading test, or simply to the fact that the students were 8 weeks older. In the case of the agronomist's experiment, changes in rainfall, temperature, or soil conditions may have been the primary reason for the healthier corn crop.

Design 3: Static Group Comparison

The static group comparison involves both an experimental group and a control group. Its design takes the following form:

Group	Time →	
Group 1	Tx	Obs
Group 2	—	Obs

An experimental group is exposed to a particular experimental treatment; the control group is not. After the treatment, both groups are observed and their performance compared. In this design, however, no attempt is made to obtain equivalent groups or even to examine the groups to determine whether they are similar before the treatment. Thus, we have no way of knowing if the treatment actually causes any observed differences between the groups.

Designs 1, 2, and 3 leave much to be desired in terms of drawing conclusions about what causes what. The experimental designs we describe next are far superior in this respect.

TRUE EXPERIMENTAL DESIGNS

In contrast with the three very simple designs just described, **experimental designs** offer a greater degree of control and, as a result, greater internal validity. The first three of the four designs we discuss in this section share one thing in common: People or other units of study are *randomly assigned to groups.* Such random assignment guarantees that any differences between the groups are probably quite small and, in any case, are due entirely to chance. The last design in this section involves a different strategy: presenting all treatments and any control conditions to a single group.

Design 4: Pretest–Posttest Control-Group Design

In a pretest–posttest control-group design, people or other units of study (e.g., members of a particular plant or animal species) are randomly assigned to either an experimental group or a control group. The experimental group is observed, subjected to the experimental treatment, and observed again. The control group is isolated from any influences of the experimental treatment; it is simply observed both at the beginning and at the end of the experiment. The basic format for the pretest–posttest control-group design is as follows:

	Group	Time →		
Random Assignment	Group 1	Obs	Tx	Obs
	Group 2	Obs	—	Obs

Such a design, simple as it is, solves two major problems associated with pre-experimental designs. We can (a) determine whether a change takes place after the treatment, and, if so, we can (b) eliminate most other possible explanations (in the form of confounding variables) as to why the change has taken place. Thus, we have a reasonable basis on which to draw a conclusion about a cause-and-effect relationship.

Design 5: Solomon Four-Group Design

One potential problem in the preceding design is that the process of observing or assessing people before administering the experimental treatment may, in and of itself, influence how people respond to the treatment. For instance, perhaps the pretest increases people's motivation: It makes them want to benefit from the treatment they receive. Such an effect is another instance of the *reactivity* effect described in Chapter 4.

To address the question *What effect does pretesting have?,* Solomon (1949) proposed an extension of the pretest–posttest control-group design that involves four groups, as depicted in the following table:

	Group	Time →		
Random Assignment	Group 1	Obs	Tx	Obs
	Group 2	Obs	—	Obs
	Group 3	—	Tx	Obs
	Group 4	—	—	Obs

The addition of two groups who are not pretested provides a particular advantage. If the researcher finds that in the final observation, Groups 3 and 4 differ in much the same way that Groups 1 and 2 do, then the researcher can more easily generalize his or her findings to situations in which no pretest has been given. In other words, the Solomon four-group design enhances the *external validity* of the study.

Compared to Design 4, this design obviously involves a larger sample and demands more of the researcher's time and energy. Its principal value is in eliminating pretest influence; when such elimination is desirable, the design is ideal.

Design 6: Posttest-Only Control-Group Design

Some life situations defy pretesting. You can't pretest the forces in a thunderstorm or a hurricane, nor can you pretest growing crops. Additionally, sometimes you may be unable to locate a suitable pretest, or, as just noted, the very act of pretesting can influence the results of the experimental manipulation. In such circumstances, the posttest-only control-group design offers a possible solution. The design may be thought of as the last two groups of the Solomon four-group design. The paradigm for the posttest-only approach is as follows:

	Group	Time →	
Random Assignment	Group 1	Tx	Obs
	Group 2	—	Obs

Random assignment to groups is critical in the posttest-only design. Without it, the researcher has nothing more than a static group comparison (Design 3), from which, for reasons previously noted, the researcher has a difficult time drawing inferences about cause and effect.

Design 7: Within-Subjects Design

Earlier we introduced you to the nature of a within-subjects design—also known as a repeated-measures design—in which all participants receive all treatments (including any control conditions) in a research study. Note that we have switched from the term *participant* to the term *subject* here. The latter term has a broader meaning than *participants* in that it can be used to refer to a wide variety of populations—perhaps human beings, dogs, or laboratory rats.

In a good within-subjects design, the various treatments are administered very close together in time, in some cases simultaneously. If we use the subscripts *a* and *b* to designate the different treatments and treatment-specific measures, then in its simplest form a within-subjects design is as follows:

Group	Time →	
Group 1	Tx_a	Obs_a
	Tx_b	Obs_b

As an example, imagine that a researcher wants to study the effects of illustrations in an instructional software program that teaches 20 science concepts to sixth graders. The software defines and describes all 20 concepts with similar precision and depth. In addition, the software illustrates 10 of those concepts (chosen randomly) with pictures or diagrams. After students have completed the software curriculum, they take a quiz that assesses their understanding of the 20 concepts, and the researcher computes separate quiz scores for the illustrated and nonillustrated concepts. If the students perform better on quiz items for illustrated concepts than on items for nonillustrated ones, the researcher can reasonably conclude that, yes, illustrations help students learn science more effectively. In other words, the researcher has identified a cause-and-effect relationship: Illustrations improve science learning.

For a within-subjects design to work, the various forms of treatment must be such that their effects are fairly localized and unlikely to "spread" beyond specifically targeted behaviors. Such is the case in the study just described: The illustrations help students learn the particular concepts that have been illustrated but don't help students learn science more generally. In contrast, it would not make sense to use a within-subjects design to study the effects of two different psychotherapeutic techniques to reduce adolescents' criminal behaviors: If the same group of adolescents receives both treatments and then shows a significant reduction in juvenile offenses, we might suspect that either treatment could have had a fairly broad impact; we wouldn't know whether one of the treatments was more effective than the other.

Ideally, too, the two different treatments should be administered repeatedly, one after another, in a balanced but somewhat random order. For example, in the instructional software that presents both illustrated and nonillustrated science concepts, we might begin with an illustrated concept, then have two nonillustrated ones, then another illustrated one, another nonillustrated one, two illustrated ones, and so on, with the presentation of the two conditions being evenly balanced throughout the program.

With the last point in mind, let's return once again to Thrailkill's dissertation involving a lecture about the American Civil War. Each group received each of the three treatments: the imagery, attention, and control conditions. The logistics of the study were such that it was difficult to intermingle the three treatments throughout the lecture; instead, Thrailkill administered first one treatment (e.g., attention), then another (e.g., imagery), and finally the third (e.g., control). Had she limited her study to a single group, she could not have ruled out an alternative explanation—*when* in the lecture the information appeared (whether it appeared near the beginning, in the middle, or at the end)—for the results she obtained. By using three different groups, each of which had any particular condition in a different part of the lecture, she was able to eliminate that alternative explanation. Strictly speaking, however, because she could neither randomize assignment to groups nor randomly distribute different treatment conditions throughout the lecture, her study is probably better characterized as a quasi-experimental study than a true experimental study. We look more closely at quasi-experimental designs now.

QUASI-EXPERIMENTAL DESIGNS

In the preceding discussion of true experimental designs, we have emphasized the importance of *randomness,* either in the selection of group members in a multiple-groups study or in the presentation of different treatments in a single-group study. Sometimes, however, randomness is either impossible or impractical. In such situations, researchers often use **quasi-experimental designs**. When they conduct quasi-experimental studies, they don't control for all confounding variables and so can't completely rule out some alternative explanations for the results they obtain. They must take whatever variables and explanations they haven't controlled for into consideration when they interpret their data.

Design 8: Nonrandomized Control-Group Pretest–Posttest Design

The nonrandomized control-group pretest–posttest design can perhaps best be described as a compromise between the static group comparison (Design 3) and the pretest–posttest control-group design (Design 4). Like Design 3, it involves two groups to which participants haven't been randomly assigned. But it incorporates the pretreatment observations of Design 4. In sum, the nonrandomized control-group pretest–posttest design can be depicted as follows:

Group	Time →		
Group 1	Obs	Tx	Obs
Group 2	Obs	—	Obs

Without random assignment, there's no guarantee that the two groups are similar in every respect prior to the experimental treatment or intervention—no guarantee that any differences between them are due entirely to chance. However, an initial observation (e.g., a pretest) can confirm that the two groups are at least similar in terms of the dependent variable under investigation. If, after one group has received the experimental treatment, we then find group differences with respect to the dependent variable, we might reasonably conclude that the posttreatment differences are probably the result of that treatment.

Identifying matched pairs in the two groups is one way of strengthening the pretest–posttest control-group design. For instance, if we are studying the effect of a particular preschool program on children's IQ scores, we might find pairs of children—each pair including one child who is enrolled in the preschool program and one who is not—who are the same age and gender and have similar IQ scores before the program begins. Although we cannot rule out all other possible explanations in this situation (e.g., it may be that the parents who enroll their children in the preschool program are, in general, more concerned about their children's cognitive development), we can at least rule out *some* alternative explanations.

Design 9: Simple Time-Series Design

In its simplest form, a time-series design consists of making a series of observations (i.e., measuring the dependent variable on several occasions), introducing an intervention or other new dynamic into the system, and then making additional observations. If a substantial change is observed in the second series of observations in comparison to the first series, we might reasonably conclude that the cause of the change was the factor introduced into the system. This design thus looks something like the following:

Group	Time →							
Group 1	Obs	Obs	Obs	Obs	Tx	Obs	Obs	Obs	Obs

In such studies, the sequence of observations made prior to the treatment is typically referred to as **baseline data**.

Such a design has been widely used in the physical and biological sciences. Sir Alexander Fleming's discovery that *Penicillium notatum* (a mold) could inhibit staphylococci (a type of bacteria) is an example of this type of design. Fleming had been observing the growth of staphylococci on a culture plate. Then, unexpectedly, a culture plate containing well-developed colonies of staphylococci was contaminated with the spores of *Penicillium notatum.* Fleming observed that the bacteria near the mold seemed to disappear. He intentionally repeated the situation: After periodically observing the bacteria, he introduced the mold. Each time he used this procedure, his subsequent observations were the same: no staph germs near the mold.

The major weakness of this design is the possibility that some other, unrecognized event in the laboratory or outside world may occur at approximately the same time that the experimental treatment does, reflecting the *history* factor described in Figure 7.1. If this other event is actually the cause of the change, any conclusion that the treatment has brought about the change will obviously be incorrect.

Design 10: Control-Group Time-Series Design

In a variation of the time-series design, two groups are observed over a period of time, but one group (a control) doesn't receive the experimental treatment. The general design takes the following form:

Group	Time →							
Group 1	Obs	Obs	Obs	Obs	Tx	Obs	Obs	Obs	Obs
Group 2	Obs	Obs	Obs	Obs	—	Obs	Obs	Obs	Obs

This design has greater internal validity than the simple time-series design (Design 9). If an outside event is the cause of any changes we observe, then presumably the performance of *both* groups will be altered after the experimental treatment takes place. If, instead, the experimental treatment is the factor that affects performance, we should see a change only for Group 1.

Design 11: Reversal Time-Series Design

The reversal time-series design uses a within-subjects approach as a way of minimizing—though not entirely eliminating—the probability that outside effects might bring about any changes observed. The intervening experimental treatment is sometimes present, sometimes absent, and we measure the dependent variable at regular intervals. Thus, we have the following design:

Group	Time →							
Group 1	Tx	Obs	—	Obs	Tx	Obs	—	Obs

To illustrate, suppose we are interested in whether audiovisual materials help a single class of students learn astronomy. On some days we might include audiovisual materials in a lesson, and on other days we might omit them. We can then measure how effectively the students learn under both conditions. If the audiovisual materials do, in fact, promote student learning, we should see consistently better student performance on those days.

Design 12: Alternating-Treatments Design

A variation on the reversal time-series design involves including two or more different forms of experimental treatment in the design. Referring to the two different forms of treatment with the notations Tx_a and Tx_b, we can depict this design in the following manner:

Group	Time →															
Group 1	Tx_a	Obs	—	Obs	Tx_b	Obs	—	Obs	Tx_a	Obs	—	Obs	Tx_b	Obs

If such a sequence were pursued over a long enough time span, we would hope to see different effects for the two different treatments.

Design 13: Multiple-Baseline Design

Designs 11 and 12 are based on the assumption that the effects of any single treatment are temporary and limited to the immediate circumstances. Thus, these designs won't work if a treatment is likely to have long-lasting and perhaps fairly general effects. Furthermore, if an experimental treatment is apt to be quite beneficial for all participants, then ethical considerations may discourage us from including an untreated control group. In such instances, a multiple-baseline design provides a good alternative. This design requires at least two groups. Prior to the treatment, baseline data are collected for all groups, and then the treatment itself is introduced at a different time for each group. In its simplest form, a multiple-baseline design might be configured as follows:

Group	Time →					
Group 1	*Baseline →*		*Treatment →*			
	—	Obs	Tx	Obs	Tx	Obs
Group 2	*Baseline →*				*Treatment →*	
	—	Obs	—	Obs	Tx	Obs

A study by Heck, Collins, and Peterson (2001) provides an example of this approach. The researchers wanted to determine if instruction in playground safety would decrease elementary school children's risky behaviors on the playground. The treatment in this case involved a 5-day intervention in which a woman visited children's classrooms to talk about potentially risky behaviors on slides and climbing equipment, as well as about the unpleasant consequences that might result from such behaviors. The woman visited four different grade levels over a 3-week period; a random selection process resulted in her visiting first-grade classes one week, second-grade classes the following week, and kindergarten and third-grade classes (which went to recess at the same time) the week after that. Meanwhile, two independent observers simultaneously

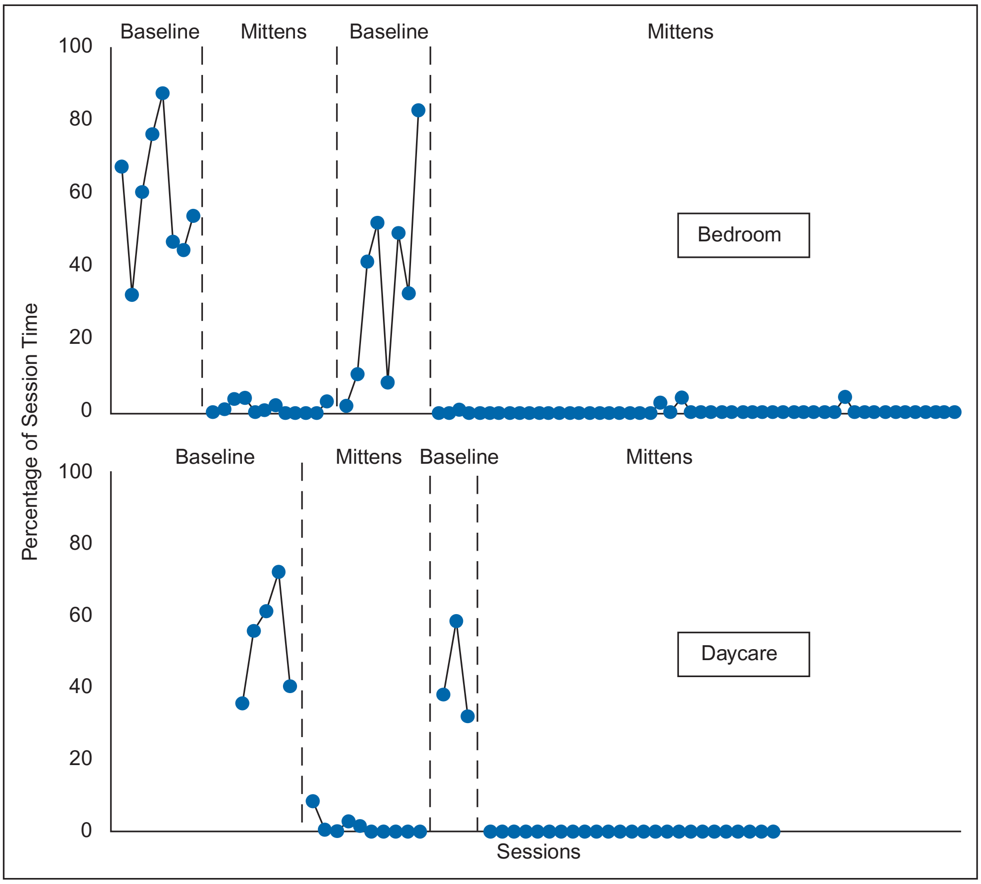
Using Designs 11, 12, and 13 in Single-Subject Studies

Reversal, alternating-treatments, and multiple-baseline designs can be used not only with groups but also with single individuals, in what are collectively known as **single-subject designs**. A study by Deaver, Miltenberger, and Stricker (2001) illustrates how a researcher might use two of these—reversal and multiple-baseline—simultaneously. A 2-year-old girl named Tina had been referred for treatment because she often twirled her hair with her fingers so vigorously that she pulled out some of her hair. On one occasion she wrapped her hair around a finger so tightly that the finger began to turn blue and the hair had to be removed with scissors. Tina engaged in such behavior primarily when she was alone (e.g., at naptime); hence, there was no parent or other adult present to discourage it. The researchers identified a simple treatment—putting thin cotton mittens on Tina's hands—and wanted to document its effect. They videotaped Tina's behaviors when she was lying down for a nap in either of two settings, her bedroom at home or her daycare center, and two observers independently counted the number of hair twirling incidents as they watched the videotapes. Initially, the observers collected baseline data. Then, during separate time periods for the bedroom and daycare settings, they gave Tina the mittens to wear during naptime. After reversing back to baseline in both settings, they had Tina wear the mittens once again. The percentages of time that Tina twirled her hair in the two settings over the course of the study are presented in Figure 7.3.

In both the bedroom and daycare settings, the researchers alternated between baseline and treatment; this is the *reversal* aspect of the study. Furthermore, they initiated and then later

FIGURE 7.3 Percentage of Session Time in Which Hair Twirling Was Observed Both in the Bedroom and at Daycare

Reprinted from "Functional Analysis and Treatment of Hair Twirling in a Young Child" by C. M. Deaver, R. G. Miltenberger, & J. M. Stricker, 2001, *Journal of Applied Behavior Analysis, 34*, p. 537. Reprinted with permission of the Society for the Experimental Analysis of Behavior, Inc.



reinstituted the treatment at different times in the two settings; this is the *multiple-baseline* aspect of the study. Figure 7.3 consistently shows dramatic differences in hair twirling during baseline versus mittens conditions, leading us to conclude that the mittens, rather than some other factor, were almost certainly the reason for the disappearance of hair twirling.

EX POST FACTO DESIGNS

In many situations, it is unethical or impossible to manipulate certain variables in order to investigate their potential influence on other variables. For example, a researcher cannot intentionally infect people with a potentially lethal new virus, withhold instruction, ask parents to abuse their children, or modify a person's personality to compare the effects of these factors on the dependent variables in one's research problem.

Ex post facto designs[2] (the term *ex post facto* literally means "after the fact") provide an alternative means by which a researcher can investigate the extent to which specific independent variables—perhaps involving a virus, lack of schooling, a history of family violence, or a personality trait—may possibly affect the dependent variable(s) of interest. In an ex post facto study, a researcher identifies *events that have already occurred or conditions that are already present* and then collects data to investigate a possible relationship between these factors and subsequent characteristics or behaviors. In particular, after observing that differing circumstances have prevailed for two or more different groups—such circumstances comprise the independent variable—the researcher tries to determine whether the groups differ on some other, dependent variable. For example, a researcher might identify two groups of adults with different immunization records—those who, as children, were vaccinated against measles and those who were not—and then calculate the percentage of reported cases of measles in each group. Similarly, a researcher might identify two groups of 10-year-olds—those who had extensive musical training in preschool and those whose preschools provided no such training—and compare the musical skills of the two groups of children.

Ex post facto designs are often confused with correlational or experimental designs because they share certain characteristics with each of these other design types. Like correlational research, ex post facto research involves looking at existing circumstances. But like experimental research, it has clearly identifiable independent and dependent variables.

Unlike experimental studies, however, ex post facto designs involve no direct manipulation of the independent variable: The presumed "cause" has already occurred. To the extent that such manipulation isn't possible, the researcher cannot draw firm conclusions about cause and effect. The problem here is that the experimenter can't control for confounding variables that might provide alternative explanations for any group differences observed.

Although an ex post facto study lacks the control element—and thus doesn't enable definite conclusions about cause and effect—it is nevertheless a legitimate research method. Medicine uses it widely in its research activities. Physicians discover an illness and then initiate their search *after the fact.* They delve into antecedent events and conditions to discover a possible cause for the illness. Such was the approach of medical researchers when the AIDS virus came to light in the 1980s.

Like experimental designs, ex post facto designs can take a variety of forms. Here we present one possible design for illustrative purposes. We present a second ex post facto design in the subsequent section on factorial designs.

[2]Ex post facto designs are also known as *causal-comparative* designs. However, as B. Johnson (2001) has pointed out, the latter term may mislead novice researchers to believe that such designs show cause and effect as clearly and definitively as true experimental designs. In reality, such designs never eliminate all other possible explanations for an observed effect; thus, they can't truly show cause and effect.

Design 14: Simple Ex Post Facto Design

Design 14 is similar to the static group comparison (Design 3) described in the previous section on pre-experimental designs. The important difference is one of timing: In this case, the "treatment" in question occurred long before the study began; hence, we call it an *experience* rather than a treatment because the researcher hasn't been responsible for imposing it. A simple ex post facto design can be depicted as follows, where Exp refers to a prior experience that one group has had and another has not:

Group	Time →	
	Prior event(s)	*Investigation period*
Group 1	Exp	Obs
Group 2	—	Obs

An obvious variation on this design is one in which Group 2 has an experience as well, albeit a different experience from that of Group 1.

Such designs are common in studying the possible effects of previously occurring environmental variables such as television viewing habits, child abuse, and malnutrition. They are also used in studying the potential influences of pre-existing characteristics—perhaps those that are inherited or congenital—such as gender, mental illness, and physical disability. (In the latter instances, we might want to replace the term *experience* with a term such as *characteristic.*) The most we can conclude from these studies is that certain behaviors or other variables tend to be *associated* with certain pre-existing conditions; we can never determine that those other variables were actually caused by those conditions.

FACTORIAL DESIGNS

Thus far we have been describing designs in which only one independent variable is studied. Yet in many situations, a researcher examines the effects of two or more independent variables in a single study; this approach is known as a **factorial design**.

Design 15: Two-Factor Experimental Design

In its simplest form—one involving two independent variables, which we will call *Variable 1* and *Variable 2*—such a design might look like the following:

	Group	Time →		
		Treatments related to the two variables may occur simultaneously or sequentially		
		Treatment related to Variable 1	*Treatment related to Variable 2*	
Random Assignment	Group 1	Tx_1	Tx_2	Obs
	Group 2	Tx_1	—	Obs
	Group 3	—	Tx_2	Obs
	Group 4	—	—	Obs

We can determine the effects of the first independent variable by comparing the performance of Groups 1 and 2 with that of Groups 3 and 4. We can determine the effects of the second independent variable by comparing Groups 1 and 3 with Groups 2 and 4. If you think you've seen this design before, in a way you have. This is simply a more generalized form of the Solomon four-group design (Design 5), but we are no longer limiting ourselves to having the presence or absence of a pretest be one of our independent variables.

Such a design allows us to examine not only the possible effects of two independent variables but also the possible *interaction* of the variables as they influence the dependent variable. For example, imagine that, after presenting both treatments, we find that Groups 2, 3, and 4 show similar performance but that Group 1 outperforms the other three. Such a result might indicate that neither independent variable produces a particular effect on its own—that *both* variables are necessary to bring about the effect.

Design 16: Combined Experimental and Ex Post Facto Design

In the factorial design just presented, participants are randomly assigned to groups in a true experimental study. But it is also possible to combine elements of experimental research and ex post facto research into a single factorial design. In its simplest form, such a design looks like the following:

<table>
<tr><th>Group</th><th colspan="5">Time →</th></tr>
<tr><td></td><td>Prior event(s)</td><td colspan="4">Investigation period →</td></tr>
<tr><td rowspan="2">Group 1</td><td rowspan="2">Exp_a</td><td rowspan="2">Random Assign-ment</td><td>Group 1a</td><td>Tx_a</td><td>Obs</td></tr>
<tr><td>Group 1b</td><td>Tx_b</td><td>Obs</td></tr>
<tr><td rowspan="2">Group 2</td><td rowspan="2">Exp_b</td><td rowspan="2">Random Assign-ment</td><td>Group 2a</td><td>Tx_a</td><td>Obs</td></tr>
<tr><td>Group 2b</td><td>Tx_b</td><td>Obs</td></tr>
</table>

In this case, the researcher initially divides the sample into two groups based on the participants' previous experiences or pre-existing conditions; this is the *ex post facto* part of the study. Then the researcher randomly assigns members of each group to one of two treatment groups (or perhaps a treatment group and a control group); this is the *experimental* part of the study. The result is four groups that represent all four possible combinations of the previous experience/pre-existing characteristic and the treatment variable. Such a design enables the researcher to study how an experimental manipulation might influence a particular dependent variable *and* how a previous experience or pre-existing characteristic might interact with that manipulation.

In a variation of such a design, the experimental manipulation might be a within-subjects variable rather than a between-groups variable. As an example, one of us authors once joined forces with two colleagues and a graduate student to test the hypothesis that people with different educational backgrounds interpret and remember maps differently and, more specifically, that only people with a background in geography apply general principles of geography when they interpret maps (J. E. Ormrod, Ormrod, Wagner, & McCallin, 1988). We constructed two maps to test our hypothesis. One map was arranged

CONCEPTUAL ANALYSIS EXERCISE Identifying Quantitative Research Designs

As a way of reviewing the designs described in this chapter, we offer a brief pop quiz. Following are short summaries of five research studies. The studies don't necessarily fit exactly into one of the design categories presented, but each one is definitely *experimental, quasi-experimental,* or *ex post facto* in nature. Identify the type of research that each study reflects. The answers appear after the "For Further Reading" section at the end of the chapter.

1. Two researchers want to see if a particular training program is effective in teaching horses to enter a horse trailer without misbehaving in the process—that is, without rearing, trying to turn around, or in some other way resisting entry into the trailer. Five horses (Red, Penny, Shadow, Sammy, and Fancy) go through the training, with each horse beginning training on a different day. For each horse, an observer counts the number of misbehaviors every day prior to and during training, with data being collected for a time span of at least 45 days (Ferguson & Rosales-Ruiz, 2001).
2. Two researchers wonder whether an eyewitness's memory of an event is affected by questions that he or she is asked subsequent to the event. To find out, the researchers show adults a film that depicts a car accident. Each adult is then asked one of five questions (randomly selected) about the accident:
   - About how fast were the cars going when they *contacted* each other?
   - About how fast were the cars going when they *hit* each other?
   - About how fast were the cars going when they *bumped into* each other?
   - About how fast were the cars going when they *collided into* each other?
   - About how fast were the cars going when they *smashed into* each other?

   The researchers compute the average speed given in response to each of the five questions to determine whether the questions have influenced participants' "memory" for the accident (Loftus & Palmer, 1974).
3. A researcher studies the effects of two different kinds of note-taking training (one of which is a placebo) on the kinds of notes that college students take. Her sample consists of students enrolled in two sections of an undergraduate course in educational psychology; with the flip of a coin, she randomly determines which section will be the treatment group and which will be the control group. She analyzes the content of students' class notes both before and after the training, making the prediction that the two groups' notes will be similar before the training but qualitatively different after the training (Jackson, 1996).
4. At the request of the National Park Service, two researchers at Rocky Mountain National Park investigate the degree to which signs along hiking trails might influence hikers' behaviors. Park Service officials are concerned that the heavy traffic on one particular hiking trail, the trail to Emerald Lake, may be having a negative impact on the local environment; they would like to divert some traffic to a lesser-used trail to Lake Haiyaha, which begins at the same place as the Emerald Lake trail. One day in early summer, the researchers hide battery-operated, optic counters at key locations along the two trails to record the number of hikers. The study has four phases: (1) at the spot where the two trails originate, only signs indicating the destinations of the two trails are present; (2) a "positively worded" sign is added that describes the attractive features of the Lake Haiyaha trail and encourages hikers to use it; (3) the positively worded sign is replaced by a "negatively worded" sign that describes the crowdedness of the Emerald Lake trail and discourages its use; and (4) both the positively worded and negatively worded signs are posted. The researchers compare the frequency of hikers during each of the four phases (R. K. Ormrod & Trahan, 1982).
5. A team of researchers has a sample of elementary school boys, some of whom have been identified as having attention-deficit hyperactivity disorder (ADHD) and some of whom have not. One of the researchers asks each boy to interpret several social situations that are depicted in a series of black-and-white drawings (e.g., one sequence

of drawings shows a sequence of events at a Halloween party). Some of the situations involve antisocial behavior (e.g., aggression), and other situations involve prosocial behavior (e.g., sharing). The researchers compare the interpretations that boys with ADHD make with the interpretations that boys without ADHD make with respect to both kinds of situations (Milch-Reich, Campbell, Pelham, Connelly, & Geva, 1999).

PRACTICAL APPLICATION Determining Possible Cause-and-Effect Relationships

The research designs described in this chapter vary considerably in the degree to which they control for potential confounding variables—variables that threaten a study's internal validity—and thus they also vary in terms of the degree to which they enable a researcher to draw firm conclusions about cause-and-effect relationships. The following checklist can help you evaluate a research design with respect to its internal validity.

CHECKLIST

Looking for Confounding Variables

If you are planning a study in which you hope to find one or more cause-and-effect relationships—or if, instead, you are evaluating another person's research proposal or report—scrutinize the study with the following questions in mind:

______ 1. What are the independent and dependent variables in the study:

Independent variable(s):

____

Dependent variable(s):

____

______ 2. Is every independent variable actively manipulated by the researcher?

______ Yes ______ No

______ 3. If the researcher is manipulating one or more independent variables, what precautions is the researcher taking to ensure that the manipulation is minimizing or eliminating the potential effects of confounding variables? For example, is the researcher:

- Keeping certain other variables constant? If so, which ones?

  ____

  ____

- Including a control group or at least two treatment groups?

  ____

- Randomizing assignment to groups?

  ____

- Using a within-subjects (repeated-measures) design?

  ____

- Using other appropriate strategies? If so, which ones?

  ____

  ____

META-ANALYSES

Remember, we can conclude that a cause-and-effect relationship exists between an independent variable and a dependent variable only if we have directly manipulated the independent variable and have controlled for confounding variables that might offer alternative explanations for any changes in the dependent variable. Even when we have taken such precautions, however, there is the possibility that our alleged "cause" doesn't really produce the effect we think it does—that the situation we have just observed is a one-time-in-a-million fluke.

In Chapter 4 we introduced the idea of *replication:* We gain greater confidence in our research findings when a study is repeated over and over again—perhaps with a different population, in a different setting, or with slight variations on the treatment implementation.

Once researchers have conducted many such replications, another researcher may come along and conduct a **meta-analysis**—that is, an analysis of the analyses. In particular, the researcher combines the results of many experimental and/or ex post facto studies to determine whether they lead to consistent conclusions. A meta-analysis is primarily a statistical technique, and thus we describe this procedure in greater depth in Chapter 8.

CONDUCTING EXPERIMENTS ON THE INTERNET

USING TECHNOLOGY

In Chapter 6 we mentioned that some researchers now conduct research studies on the Internet. Although most of these studies can best be categorized as descriptive studies, we occasionally see experimental studies as well. For instance, one of us authors once visited the website "Psychological Research on the Net," which provides links to numerous sites that host online research projects.[3] To learn more about this growing approach to data collection, she became a participant in several online studies that were active at the time. Although most of the studies involved completing questionnaires and so appeared to be correlational or survey studies, one of them was clearly an experimental study. In particular, this author was asked to (a) read and study a story that was illustrated by several photographs; (b) read three additional stories, one of which was quite similar to the initial story; and (c) answer a series of questions about details in the stories. In a subsequent debriefing on the website, she learned that she had been randomly assigned to the experimental group in the second part of the study; other participants were assigned to a control group, in which all three stories were quite different from the initial story. The researcher was investigating the possible effects that a similar story in Part b might have on recall for the story in Part a.

Internet-based experimental studies don't necessarily have to be one-shot affairs. For example, in one online study (Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008), researchers enticed people into participating in a three-session experiment with the promise that for every session they completed, their name would be entered into an end-of-study lottery that would award cash prizes. A total of 1,354 people completed all three sessions; they ranged in age from 18 to 72 and lived in various countries around the world. In Session 1 of the experiment, participants studied a list of 32 obscure trivia facts, such as the answer to "What European nation consumes the most spicy Mexican food?" (p. 1097), and they continued to study each fact until they could correctly recall it.[4] After this first session, participants were divided into different treatment groups that varied in terms of the timing for Sessions 2 and 3, and they were sent e-mail messages when it was time to complete these subsequent sessions. In Session 2 (which might be as little as 3 minutes or as much as 105 days after Session 1), participants studied the trivia facts again, this time studying each one twice. Then, in Session 3 (which was 7, 35, 70, or 350 days after Session 2), participants were asked to remember as many of the facts as they could. The findings of the study are important for any conscientious student to note: Especially when the final test session was considerably delayed

[3]As noted in Chapter 6, this website is maintained by John Krantz, Professor of Psychology at Hanover College (psych.hanover.edu).

[4]In case you're curious, Norwegians are especially partial to spicy Mexican food, at least in comparison with other Europeans.

(e.g., by 2½ months or almost a year), people who spread out their studying more (i.e., those with a longer delay between Sessions 1 and 2) remembered more facts. (If you've noticed a possible problem with *attrition* in the study, give yourself a pat on the back! We'll address this problem shortly.)

In some instances, an Internet-based research study might be quite suitable for your research question. Keep in mind, however, that ethical practices ensuring protection from harm, informed consent, and right to privacy are as important in online experimental research as they are in any face-to-face studies. The suggestions for ethical practices presented in Chapter 6 for online questionnaires are equally applicable to online experiments (see the Practical Application "Using the Internet to Collect Data for a Descriptive Study" in Chapter 6).

Remember, too, that the sample you get in an online study will hardly be representative of the overall population; for instance, it is likely to consist largely of college-educated, computer-literate people who enjoy participating in research studies. An additional problem is that you cannot observe your participants to determine whether they are accurately reporting demographic information (their age, gender, etc.) and whether they are truly following the instructions you present. Accordingly, unless you are interested in a topic such as very-long-term memory (as Cepeda and his colleagues were in their 2008 study) and can carefully control the conditions under which people are participating, we suggest that you use an Internet-based study primarily to formulate tentative hypotheses or to pilot test experimental materials you plan to use in a more controlled and observable situation.

TESTING YOUR HYPOTHESES, AND BEYOND

Experimental and ex post facto studies typically begin with specific research hypotheses, and subsequent statistical analyses should, of course, be conducted to test these hypotheses. Such analyses often take the form of a *t* test, analysis of variance, or analysis of covariance. We briefly describe these procedures in Chapter 8.

Yet one's analyses need not be restricted *only* to the testing of initially stated hypotheses. Oftentimes a study may yield additional results—results that are unexpected yet intriguing—that merit analysis. There is no reason why the researcher can't examine these findings as well, perhaps statistically, perhaps not.

PRACTICAL APPLICATION Acknowledging the Probable Presence of Bias in Experimental Research

Despite the tight controls in many experiments—and in some cases *because* of such controls—one or more forms of bias can wiggle their ways into the data or into interpretations of the data. Some of these biasing factors, such as group selection procedures, statistical regression, and differing attrition rates, can adversely affect the *internal validity* of a study (look once again at Figure 7.1). For example, as you were reading about the memory-for-trivia experiment in the earlier discussion of Internet-based experiments, you might have wondered if the dropout (attrition) rate was higher for participants with longer between-session delays, and indeed it was (Cepeda et al., 2008). Were participants who had poor memories more likely to drop out over the long run than participants who had good memories? If so, by Session 3, the people who remained in spread-out-studying treatment groups might simply have had better memories *in general* than people who remained in close-together-studying treatment groups. To determine the extent to which the differing attrition rates for various treatment groups might jeopardize the study's internal validity, the researchers collected basic demographic data at the beginning of Session 1. In their data analyses, the researchers found no significant differences in any demographic variables or in Session 1 performance between participants who completed all three sessions and those who did not—thus lending support to their premise that the members of the various treatment groups were similar in all ways *except* for the differing study intervals.

A SAMPLE DISSERTATION

To illustrate how an experimental study might appear in its written form, we present excerpts from Virginia Kinnick's doctoral dissertation conducted at the University of Colorado (Kinnick, 1989). The researcher, a faculty member in the School of Nursing at another university, had considerable experience teaching nursing students the knowledge and skills they would need when working with women who were in the process of delivering a baby, and her interest lay in learning more about teaching such knowledge and skills effectively.

During a woman's labor prior to the delivery of her baby, a fetal monitor is often used to assess the baby's heart rate, and the maternity nurse must frequently check the monitor for signs that the baby might be experiencing exceptional and potentially harmful stress. Kinnick wanted to determine whether a particular method of teaching concepts (one described by Tennyson and Cocchiarella) might be more effective for teaching fetal monitoring skills than the method traditionally used in nursing education programs. In Kinnick's dissertation, the problem statement is as follows:

> This study is designed to determine if use of an instructional design model for concept attainment in teaching the critical concepts related to fetal monitoring will make a significant difference in preparation of nursing students in this skill, compared to the traditional teaching method which exists in most schools. (Kinnick, 1989, p. 8)

The research design was not one of the designs we have specifically described in this chapter. Instead, it involved administering three different instructional methods to three treatment groups (with participants assigned randomly to groups) and then observing the effects of the treatments at two different times: once immediately after instruction and then later after students had completed the clinical rotation portion of their nursing program. Thus, the design of the study was the following:

	Group	Time →		
Random Assignment	Group 1	Tx_1	Obs	Obs
	Group 2	Tx_2	Obs	Obs
	Group 3	Tx_3	Obs	Obs

In the following pages, we present excerpts from the methodology chapter of the researcher's dissertation. Our comments and observations appear on the right-hand side.

FOR FURTHER READING

Antony, J. (2003). *Design of experiments for engineers and scientists.* Oxford, England: Butterworth-Heineman/Elsevier.

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Upper Saddle River, NJ: Pearson.

Bausell, R. B. (1994). *Conducting meaningful experiments: Forty steps to becoming a scientist.* Thousand Oaks, CA: Sage.

Canavos, G. C., & Koutrouvelis, I. A. (2009). *An introduction to the design and analysis of experiments.* Upper Saddle River, NJ: Pearson.

Dugard, P., File, P., & Todman, J. (2012). *Single-case and small-n experimental designs: A practical guide to randomization tests* (2nd ed.). New York: Routledge.

Friedman, D., & Sunder, S. (1994). *Experimental methods: A primer for economists.* New York: Cambridge University Press.

Glass, D. J. (2007). *Experimental design for biologists.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Glass, G. V. (1988). Quasi experiments: The case of interrupted time series. In R. M. Jaeger (Ed.), *Complementary methods for research in education* (pp. 445–464). Washington, DC: American Educational Research Association.

Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Thousand Oaks, CA: Sage.

Montgomery, D. C. (2012). *Design and analysis of experiments* (8th ed.). New York: Wiley.

Morgan, D. L., & Morgan, R. K. (2001). Single participant research design: Bringing science to managed care. *American Psychologist, 56,* 119–127.

O'Neill, R. E., McDonnell, J. J., Billingsley, F., & Jenson, W. (2011). *Single case research designs in educational and community settings.* Upper Saddle River, NJ: Pearson.

Phillips, D. C. (1981). Toward an evaluation of the experiment in educational contexts. *Educational Researcher, 10*(6), 13–20.

Pukelsheim, F. (2006). *Optimal design of experiments (Classics in applied mathematics).* New York: Wiley.

Ruxton, G., & Colegrave, N. (2010). *Experimental design for the life sciences* (3rd ed.). Oxford, England: Oxford University Press.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs.* Washington, DC: American Educational Research Association.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Solso, R. S., & MacLin, K. (2008). *Experimental psychology: A case approach* (8th ed.). Upper Saddle River, NJ: Pearson.

ANSWERS TO THE CONCEPTUAL ANALYSIS EXERCISE "Identifying Quantitative Research Designs":

1. This is a *quasi-experimental* study. In particular, it involves a *multiple-baseline* design: Each of the horses begins training on a different day. In the section of the chapter "Using Designs 11, 12, and 13 in Single-Subject Studies," a multiple-baseline study is described in which a single 2-year-old girl successively received a particular treatment (having mittens) in two different contexts. In this example, however, we see the approach being used with five different horses, each of which receives the treatment only once.
2. This is an *experimental* study in which the researchers randomly assign participants to one of five groups, each of which is asked a different question.
3. Don't let the random selection of treatment and control groups fool you. This is a *quasi-experimental* study because participants are not randomly assigned as *individuals* to the treatment and control groups. More specifically, the study is a *nonrandomized control-group pretest–posttest* design (Design 8).
4. This, too, is a *quasi-experimental* study. It is a *time-series* design in which the effects of no intervention (Phase 1) are compared to the effects of two different interventions (the two new signs) imposed either singly or in combination. Of the designs described in this chapter, it is probably most similar to Design 12. Note, however, that no phase of the study is repeated; this omission is a decided weakness in the design.
5. This is a *combined experimental and ex post facto factorial* design with two independent variables, one of which is a *within-subjects* variable. One independent variable is the presence or absence of attention-deficit hyperactivity disorder, which the researchers do not (and *cannot*) manipulate; this is the ex post facto component of the design. The other independent variable is the content of the drawings (aggression vs. prosocial behavior); this is the experimental, within-subjects component of the design.