This item is provided to support UOB courses.

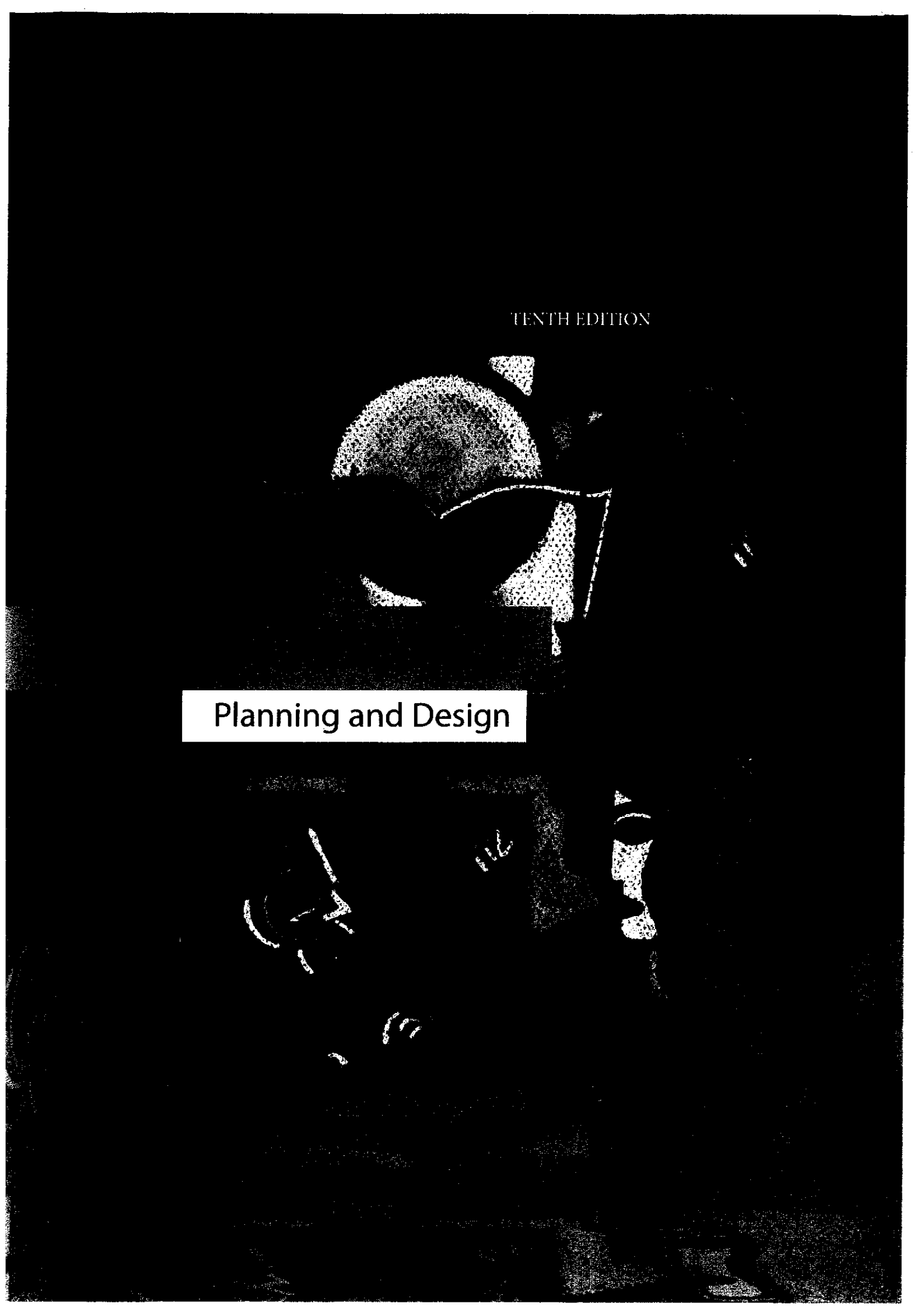
Its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission.

However, users may print, download, or email it for individual use for learning and research purposes only.

هذه الوثيقة متوفرة لمساندة مقرارات الجامعة.

ويمنع منعا باتاَ نسخها في نسخ متعددة أو إرسالها بالبريد الالكتروني الى قائمة تعميم بدون الحصول على إذن مسبق من صاحب الحق القانوني للملكية الفكرية لكن يمكن للمستفيد أن يطبع أو يحفظ نسخة منها لاستخدام الشخصي لأغراض التعلم والبحث العلمي فقط.

TENTH EDITION



Planning and Design

Pearson's MyLab™

PROVEN RESULTS

For over 10 years, instructors and students have reported better grades through increased engagement and real-time insights into progress.

ENGAGING EXPERIENCES

MyLab is designed to reach students in a personal way. Engaging learning and practice opportunities lead to assessments that create a personalized study plan.

A TRUSTED PARTNERSHIP

With millions of students registered annually, MyLab is the most effective and reliable learning solution available today.

To order this text with **MyEducationLab™**
use ISBN 0-13-289361-4

ISBN-13: 978-0-13-269324-0
ISBN-10: 0-13-269324-0

EAN 9 780132 693240 90000 >

www.pearsonhighered.com

PEARSON

ALWAYS LEARNING

9 Experimental, Quasi-Experimental, and Ex Post Facto Designs

Progress is relative: We measure it by noting the degree 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 causal effect on the phenomenon we are studying. To the extent that many potentially causal factors all vary at once, in a confounded manner, we learn little or nothing about what causes what.

In the designs described in preceding chapters, we have made no systematic attempt to determine the causes of the phenomena being studied. But ultimately we often 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 not only observed but also *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 introduced you to independent and dependent variables in Chapter 2, but because these concepts will guide so much of our discussion in this chapter, a brief refresher might be in order here. An **independent variable** is one that the researcher studies as a possible cause of something else. In many of the designs described in this chapter, the researcher directly manipulates the independent variable. In contrast, a **dependent variable** is a variable that is potentially influenced by the independent variable—it is the "something else" the independent variable

possibly influences, and thus it *depends* on the independent variable. In other words, the hypothesized relationship is this:

Independent variable → Dependent variable

As an example, let's look at a dissertation in educational psychology (Thrailkill, 1996). The researcher 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: (1) she described certain historical figures and events in such a manner that they were easy to imagine and visualize (*imagery* condition), (2) she included attention-grabbing phrases in the lecture (*attention* condition), or (c) she did neither of these things (*control* condition). In the following examples, 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 the dimpled, baby-faced young blond 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. LISTEN TO ME NOW. 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. Lincolw the less experienced junior officer 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, the researcher 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 the 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 is 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 science 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, teaching styles, and so on. 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 cannot pin down *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 classic book chapter, Campbell and Stanley (1963) identified several potential threats to the internal validity of an experimental study; we describe them in Figure 9.1.

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 sex, age, 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 Chapter 4's discussion of this concept).

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 two-month period, even just the knowledge that an experiment is being conducted—that may also account for the typists' 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.

As you should recall from Chapter 4, people sometimes show improved performance simply because they know they are participating in a research study—a phenomenon known as *reactivity* and, more specifically, the *Hawthorne effect*. To take this fact into account, a researcher sometimes gives the people in a control group a **placebo** that has the appearance of having an effect but in reality *shouldn't* have an effect. For instance, a researcher studying the effects of a new arthritis medication might give some participants a particular dosage of the medicine and give others a

FIGURE 9.1
Potential threats to the internal validity in an experimental study

When a researcher studies the possible effects of an intervention on some other (dependent) variable, a number of confounding variables can come into play that threaten the study's internal validity and thereby also jeopardize any cause-and-effect conclusions the researcher might hope to draw. Campbell and Stanley (1963) have identified the following potential threats to internal validity, which can be present either singly or in combination:

1. *History:* An uncontrolled outside event occurring between two measurements of the dependent variable brings about a change in the dependent variable. For example, a noteworthy event in the local community might change participants' knowledge, abilities, or emotional states in ways that affect the second measurement of the dependent variable.
2. *Maturation:* A change in participants' characteristics or abilities might simply be the result of the passage of time. For example, children might make normal developmental gains in eye-hand coordination or intellectual ability.
3. *Testing:* Taking a test at one time influences participants' performance during a subsequent administration of the test, perhaps simply as a result of practice in taking the test. For example, people who take a multiple-choice test at one time may gain general test-taking skills that enhance their performance on a subsequent multiple-choice test.
4. *Instrumentation:* A change occurs in how a measurement instrument is used from one time to the next. For example, a researcher might have one research assistant rate participants' performance on the first occasion but have a different research assistant judge their performance on the subsequent occasion. Any observed change might be the result of the two assistants' differing standards for rating the performance. (This threat to internal validity reflects a problem with interrater reliability; see Chapter 4).
5. *Statistical regression:* People who score extremely high or low on a measure at one time are likely to score in a less extreme manner on the same measure at a later time; that is, extreme scorers tend to "drift" toward more average performance during a subsequent measure. For example, a researcher might assign people to one of two groups—"high-anxiety" or "low-anxiety"—based on their extremely high or low scores on a self-report questionnaire designed to measure general anxiety level. Especially if the initially extreme scores were the result of people's temporary circumstances—circumstances that might make them feel either exceptionally anxious or, instead, quite "mellow" on the first testing—the supposedly high-anxiety people would become less anxious and the supposedly low-anxiety people would become more anxious regardless of any experimental interventions the two groups might undergo.
6. *Selection:* A bias exists in how members of different groups in a study are chosen. For example, when recruiting college students for a study, a researcher might put all students enrolled in an 8:00 a.m. class in one treatment group and all students enrolled in a 2:00 p.m. class in another treatment group. Students taking the early-morning class might be different in some significant way from those taking the afternoon class (e.g., the sleeping habits of the two groups might be different).
7. *Attrition:** Members of different groups drop out of the study at proportionally different rates. For example, one group in a study might lose 25% of its members before the final measurement, whereas another group might lose only 5% of its members. Thus, even if the two groups were equivalent with regard to important characteristics at the beginning of the study, they might be different in some significant way later in the study simply as a result of the differential dropout rate.

Campbell and Stanley listed an eighth threat to internal validity as well: an *interaction* among two of the threats listed above. For example, if students in an 8:00 a.m. class are assigned to one treatment group and students in a 2:00 p.m. class are assigned to a different treatment group, and if students in the 8:00 a.m. group drop out of the study in greater numbers than students in the 2:00 p.m. group, any final differences observed in the dependent variable might be the result of the fact that early risers are, for some reason, more likely to drop out than students who like to sleep in a bit. In this situation, it becomes virtually impossible to disentangle possible effects of an experimental intervention from effects of (a) the selection bias, (b) the differing dropout rates, and (c) the interaction of these two confounding variables.

**Note:* Campbell and Stanley use the term *experimental mortality* for this threat to internal validity, but the term *attrition* is more commonly seen in contemporary research literature.

similar-looking sugar pill. Or a researcher investigating a new approach to treating test anxiety might use the new treatment with some individuals but give other individuals general relaxation training that, although possibly beneficial in other ways, won't necessarily address their test anxiety.

We must emphasize—and we emphasize it quite strongly—that any researcher who incorporates placebos in a study must consider *three ethical issues* related to the use of placebos. First is the principle of informed consent: Participants in the study must be told that the study includes a placebo treatment as well as an experimental treatment and that they won't know which treatment they have received until the study has ended. Second, if participants in the study have actively sought help for a medical, psychological, or other significant problem, those who initially receive the placebo treatment should, at the conclusion of the study, be given the opportunity to receive more effective treatment. (This is assuming, of course, that the treatment *is* more effective than the placebo.) Third, and most important, when studying a treatment related to life-threatening situations (e.g., a new drug for terminal cancer, a new psychotherapeutic technique for suicidal teenagers), the researcher must seriously weigh (a) the benefits of the new knowledge that can be gained by a control group receiving no treatment against (b) the lives that may be saved by including all participants in the treatment group.

Our last point raises an issue we cannot possibly resolve for you here. Should you find yourself having to make a decision about the best research design to use in a life-and-death situation, you should consult with your professional colleagues, the internal review board at your institution, and, of course, your own conscience.

3. *Randomly assign people to groups.* In Chapter 8 we spoke at length of the value of selecting people at random to participate in a descriptive research study; such random selection enhances the probability that any results obtained for the sample also apply to the population from which the sample has been drawn. In experimental studies, researchers use random selection for a different purpose: to assign participants within their sample to various groups.

In any research study involving human beings or other living things, members of the sample are likely to be different from one another in many ways that are relevant to the variables under investigation. For example, earlier in the chapter we described a situation in which a researcher wants to compare two methods of teaching elementary school science. The students in the study will almost certainly differ from one another in intelligence, motivation, educational opportunities at home, and other factors that will affect their performance on the achievement test given at the end of the school year. It would be virtually impossible to control for such variables by having all students in the study have the *same* intelligence, the *same* motivation, the *same* kinds of outside opportunities, and so on.

As an alternative to keeping some characteristics the same for everyone, a researcher can, instead, randomly assign participants to groups. When people have been selected for one group or another on a random basis, the researcher can reasonably assume that, *on average, the groups are quite similar* and that *any differences between them are due entirely to chance.* In fact, many inferential statistical tests—especially those that allow the researcher to make comparisons among two or more groups—are based on the assumption that group membership is randomly determined and that any pretreatment differences between the groups result from chance alone.

4. *Assess equivalence before the treatment with one or more pretests.* Sometimes random assignment to two different groups simply isn't possible; for instance, researchers may have to study groups that already exist (e.g., students in classrooms, participants in different medical treatment programs). An alternative in this situation is to assess other variables that might influence the dependent variable and determine whether the groups are similar with respect to those variables. If the groups *are* similar, the probability that such variables could account for any future group differences is reduced considerably.

Another strategy is to identify **matched** pairs: pairs of people—one in each of two groups being compared—who are identical or very similar with respect to characteristics that may potentially have an effect on the dependent varaible. For instance, a researcher comparing the achievement test scores of students in two different instructional programs might identify pairs

of students of the same sex and age who have similar IQ scores. A researcher comparing two different treatments for a particular illness might match patients according to sex, age, and duration and intensity of the illness. In either case, the researcher does not study the data collected for *all* people in the two groups, only the people who are part of "matched sets" that he or she has identified. A researcher who uses this approach will, in the final research report, explain in what way(s) the participants in the study have been matched. For example, he or she might say, "Pairs were matched on the basis of sex, age, and IQ."

One problem with assessing before-treatment equivalence with pretests is that the researcher rules out *only the variables that he or she has actually assessed and determined to be equivalent across groups.* The design does not rule other influential factors that the researcher has not assessed and perhaps not even considered.

5. *Expose participants to all experimental conditions.* Still another strategy for controlling for individual differences is to *use participants as their own controls*—that is, to have every participant in the study undergo all experimental and control treatments and then assess the effects of each treatment independently. Any independent variable that is systematically varied for every participant is known as a **within-subjects variable**, and an approach that includes a within-subjects variable is known as a **within-subjects design**. You may also see the term **repeated-measures design** used in reference to this approach.

As an example, let's return to the dissertation involving three different lecture methods and their possible effects on recall for lecture content (Thrailkill, 1996). The researcher's sample consisted of volunteer students who were enrolled in three sections of an undergraduate class in educational psychology, and she planned to give the lecture just three times, once to each class. The lecture was about an American Civil War battle sufficiently obscure that participants were unlikely to have had any prior knowledge about it; thus, participants' prior knowledge about the battle was a constant—they all had *zero* prior knowledge—rather than a confounding variable. The researcher divided the lecture into three parts of approximately equal length and wrote three versions of each part, one version each for the imagery, attention, and control conditions. She combined the three versions of the three lecture parts such that each class received the different treatments in a different sequence, as follows:

	PART OF LECTURE		
	First Part	*Middle Part*	*Last Part*
Group 1	Attention	Imagery	Control
Group 2	Control	Attention	Imagery
Group 3	Imagery	Control	Attention

In this manner, all participants in her study were exposed to the two treatments and the control condition, and each condition occurred in all possible places (first, second, and third) in the sequence.

In the study just described, the researcher used a within-subjects variable (type of intervention: imagery vs. attention vs. control) to compensate for the fact that participants had not been randomly assigned to the three class sections in her sample. Sometimes researchers use a similar strategy with just a single group, and in some cases with just a single individual. You will learn some strategies for showing causation in single-group and single-individual studies later in the chapter, when we explore *quasi-experimental designs.*

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* are suitable for

this purpose. We briefly describe each of these in Chapter 11. 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 relationships.*

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 is 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 a starting point that gets you thinking about how you might best tackle your own research problem.

[1]In particular, Designs 1–6 and Designs 8–11 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 is 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 boy sitting on the damp ground in mid-April. The next day he has a sore throat and a cold. We conclude that sitting on the damp earth caused him 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 door brings good fortune to the house. 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 case study design of many qualitative studies. As described in Chapter 6, 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 eight weeks, and then administers an alternate form of the same standardized reading test. If the students' test scores improve over the eight-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 hybridizes two strains of corn. She finds that the hybrid strain is more disease-resistant and has a better yield than either of the two parent types. She concludes that the hybridization 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 have not 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 eight 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.

The three designs just described 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, an experimental group and a control group are carefully selected through appropriate randomization procedures. 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 that are not pretested provides a distinct 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 cannot pretest the forces in a thunderstorm or a hurricane, nor can you pretest growing crops. Additionally, at times 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, of course, 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, pigeons, 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 teaching science concepts to sixth graders. The researcher creates a short textbook that presents, say, 20 different concepts. In the text, all 20 concepts are defined and described with similar precision and depth. In addition, the text illustrates 10 of those concepts (chosen randomly) with pictures or diagrams. After students read the text, 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. This 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.

Ideally, too, the two different treatments should be administered repeatedly, one after another, in a balanced but somewhat random order. For example, in the textbook 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 book.

With the last point in mind, let's return to the dissertation involving the American Civil War lecture described earlier. 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, the researcher administered first one treatment (e.g., attention), then another (e.g., imagery), and finally the third (e.g., control). Had the researcher 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 the researcher 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 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 cannot completely rule out some alternative explanations for the results they obtain. They must take whatever variables and explanations they have not 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 have not 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 is, of course, 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 sex and age 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 9.1. If this other event is actually the cause of the change, any conclusion that the treatment has brought about the change will, of course, be an erroneous one.

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 design is configured as follows:

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 8). 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, then 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 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 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 way:

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 apt to have long-lasting and perhaps fairly general effects. Furthermore, if an experimental treatment is likely 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 →					
	Baseline →		*Treatment* →			
Group 1	—	Obs	Tx	Obs	Tx	Obs
	Baseline →				*Treatment* →	
Group 2	—	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 five-day intervention in which a woman visited the classroom 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 classrooms on different weeks; a random selection process resulted in her visiting the first-grade class one week, the second-grade class the following week, and the kindergarten and third-grade classes (which went to recess at the same time) the week after that. Meanwhile, two independent observers simultaneously counted the number of risky behaviors on the playground before, during, and after the interventions in

the four classrooms. The data they collected are depicted in Figure 9.2; numbers of risky behaviors on the slide are shown with the lighter dots, whereas those on the climbing equipment are shown with the darker dots. Notice that each group has data for three time periods: a pre-intervention baseline period, the five-day safety-training period, and a posttraining follow-up period. As you can see, once safety training began, the children in the second-grade and third-grade/kindergarten groups showed noticeable declines in risky behaviors on the slide and, to a lesser extent, on the climbing equipment (where risky behavior was relatively infrequently to begin with). Because the behavior changes occurred at different times for these two groups, and in particular because the changes for each group occurred at about the time that the group began its safety training, the researchers reasonably concluded that the training itself (rather than some other factor) was probably the reason for the changes. The first graders, who received the training first, showed little or no benefit from it, especially for the climbing equipment. Perhaps the trainer was still perfecting her training procedures that first week; however, we have no way of knowing for sure why the training appeared to be relatively ineffective for the first group.

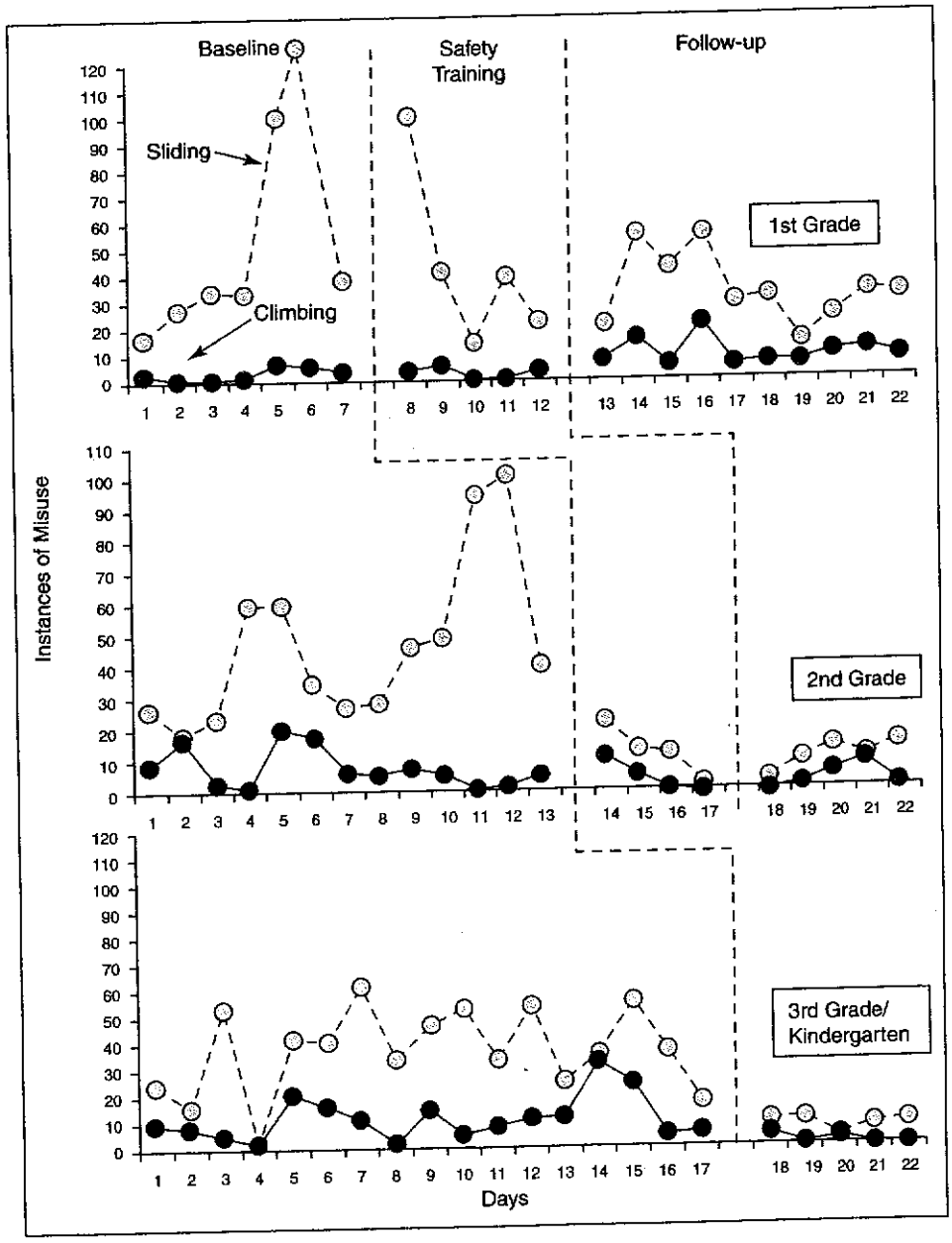


FIGURE 9.2

Instances of risky behavior on slides and climbers by grade level; third graders and kindergartners shared a single recess

Reprinted from "Decreasing Children's Risk Taking on the Playground" by A. Heck, J. Collins, and L. Peterson, 2001, *Journal of Applied Behavior Analysis, 34,* p. 351. Reprinted with permission of the Society for the Experimental Analysis of Behavior, Inc.

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 9.3.

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 reinstituted the treatment at different times in the two settings; this is the *multiple baseline* aspect of the study. Figure 9.3 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.

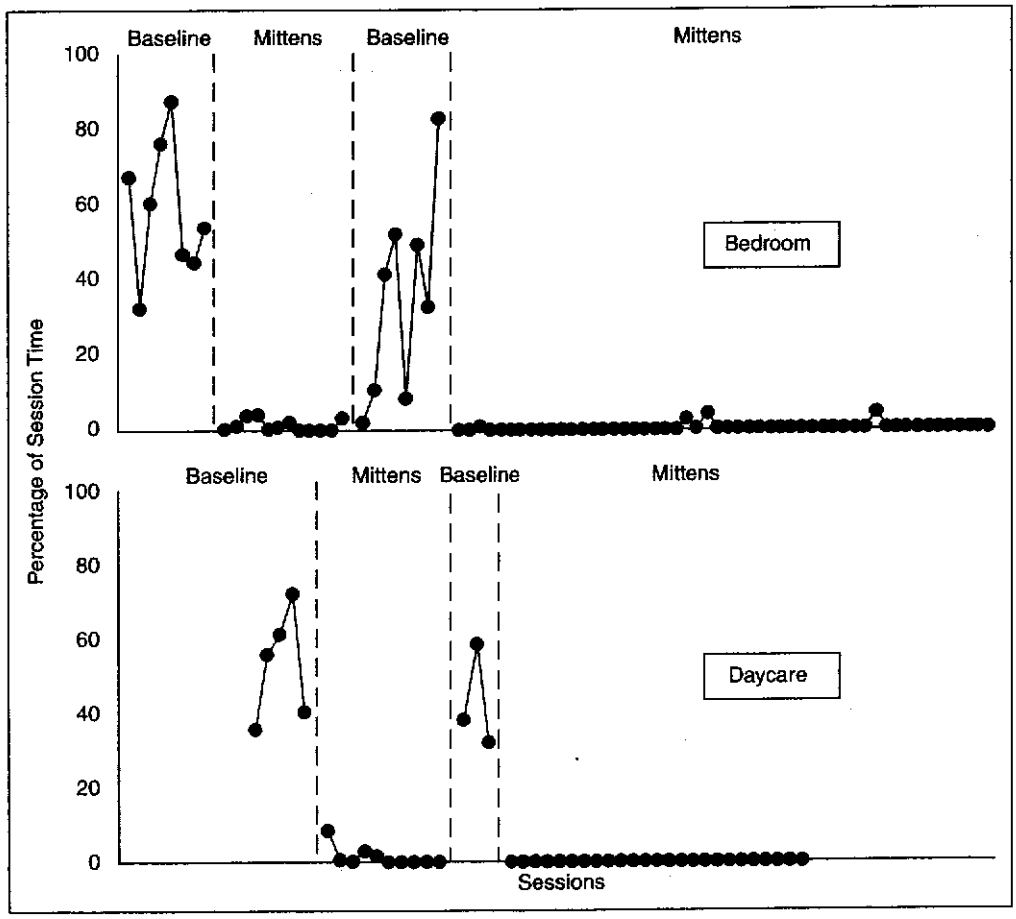


FIGURE 9.3

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.

Ex Post Facto Designs

In many situations, it is either 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 deadly 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 is not possible, the researcher cannot draw firm conclusions about cause and effect. The problem here is that the experimenter cannot control for confounding variables that may provide alternative explanations for any group differences that are observed.

Although an ex post facto study lacks the control element—and so does not allow us to draw definite conclusions about cause and effect—it is nevertheless a legitimate research method that pursues truth and seeks the solution of a problem through the analysis of data. 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 emerged 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.

Design 14: Simple Ex Post Facto Design

Design 14 is similar to the static group comparison (Design 3), which we included in our discussion of 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 has not been responsible for imposing it. A simple

[2] Ex post facto designs are also known as *causal-comparative* designs. However, as 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 cannot truly show cause and effect.

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 call *Variable 1* and *Variable 2*—such a design might look something 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.

river originated in the plains and ran *up* into a mountain range, and various transportation networks did not interconnect in ways that they normally do. The two different maps reflected one of our independent variables: logic (or lack thereof) of the spatial arrangement of features within a map.

Three groups of college professors—geographers, sociologists, and educational psychologists—provided the basis for our second independent variable: educational background. We asked each professor to study each of the two maps aloud for three two-minute intervals (we tape-recorded what they said during the study sessions) and then, after each interval, to draw as much of the map as he or she could remember.

Thus, if we call the two maps Tx_a (logical map) and Tx_b (illogical map), our design looked like the following:

Group	Time →							
Geographers	Tx_a	Obs	Obs	Obs	Tx_b	Obs	Obs	Obs
Sociologists	Tx_a	Obs	Obs	Obs	Tx_b	Obs	Obs	Obs
Educational psychologists	Tx_a	Obs	Obs	Obs	Tx_b	Obs	Obs	Obs

In this situation, one independent variable—the logic or illogic of the map presented—was a variable we directly manipulated, and we presented it to all participants in a *within-subjects* (repeated-measures) manner. The second independent variable, educational background, was a pre-existing condition and therefore something we could *not* control; this was the *ex post facto* part of the design.

The upshot of the study was that there was an *interaction* between the two independent variables, map logic and educational background. In particular, the geographers remembered more of the logical map than they did of the illogical map; in contrast, the sociologists and educational psychologists remembered each map with equal accuracy. We interpreted this result to indicate that only the geographers were applying geographic principles to study the maps and that they could use such principles effectively only with the geographically logical one. We supported our conclusion with a qualitative element in our study; that is, we used a *mixed-methods design.* In particular, we conducted content analyses of the professors' study sessions. Indeed, the content analyses revealed that the geographers had applied many geographic principles to the logical map but had trouble applying them to the illogical one. Meanwhile, the sociologists and educational psychologists studied both maps in a haphazard fashion, with few attempts to make sense of what they saw on the maps.

Table 9.1 provides a summary of the pre-experimental, experimental, quasi-experimental, ex post facto, and factorial designs described in the preceding sections. Keep in mind that, as stated earlier, this is not an exhaustive list of experimental and ex post facto designs. You can combine and expand on these designs in a number of ways—and perhaps incorporate elements of qualitative or descriptive-quantitative designs (e.g., content analysis or longitudinal data collection) as well—to more effectively address your own research question.

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 may 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:

Group	Time →				
	Prior event(s)	*Investigation period* →			
Group 1	Exp_a	Random Assign-ment	Group 1a	Tx_a	Obs
			Group 1b	Tx_b	Obs
Group 2	Exp_b	Random Assign-ment	Group 2a	Tx_a	Obs
			Group 2b	Tx_b	Obs

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 may influence a particular dependent variable *and* how a previous experience or pre-existing characteristic may possibly 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 in accordance with the patterns of a typical city; for instance, a downtown business district was located at a point where it could be easily reached from different directions (this is typical), and factories, a lumberyard, and low-income housing were situated near railroad tracks (also typical). The second map was less "logical" in the sense that it violated basic geographic principles; for instance, a

TABLE 9.1 Summary of Experimental and Ex Post Facto Designs *(continued)*

		Graphic Depiction	Comments on the Design
		Quasi-Experimental Designs *(continued)*	
9. Simple time-series experiment	To show that, for a single group, change occurs during a lengthy period only after the treatment has been administered	Group Time → Group 1: Obs, Obs, Tx, Obs, Obs	Provides a stronger alternative to Design 2; external validity can be increased by repeating the experiment in different places under different conditions.
10. Control group, time-series design	To bolster the internal validity of the preceding design with the addition of a control group	Group Time → Group 1: Obs, Obs, Tx, Obs, Obs Group 2: Obs, Obs, —, Obs, Obs	Involves conducting parallel series of observations for experimental and control groups.
11. Reversal time-series design	To show, in a single group or individual, that a treatment consistently leads to a particular effect	Group Time → Group 1: Tx, Obs, —, Obs, Tx, Obs	Is an on-again, off-again design in which the experimental treatment is sometimes present, sometimes absent.
12. Alternating treatments design	To show, in a single group or individual, that different treatments have different effects	Group Time → Group 1: Tx_a, Obs, —, Obs, Tx_b, Obs	Involves sequentially administering different treatments at different times and comparing their effects against the possible consequences of nontreatment.
13. Multiple baseline design	To show the effect of a treatment by initiating it at different times for different groups or individuals, or perhaps in different settings for a single individual	Group Time → Group 1: —, Obs, Tx, Obs, Tx, Obs Group 2: —, Obs, —, Obs, Tx, Obs	Involves tracking two or more groups or individuals over time, or tracking a single individual in two or more settings, for a lengthy period of time, as well as initiating the treatment at different times for different groups, individuals, or settings.
		Ex Post Facto Designs	
14. Simple ex post facto design	To show the possible effects of an experience that occurred, or a condition that was present, prior to the investigation	Group Time → Group 1: Exp, Obs Group 2: —, Obs	May show a difference between groups but does not conclusively demonstrate that the difference is due to the prior experience/condition in question.
		Factorial Designs	
15. Two-factor experimental design	To study the effects of two experimenter-manipulated variables and their possible interaction	Group Time → Random Assignment Group 1: Tx_1, Tx_2, Obs Group 2: Tx_1, —, Obs Group 3: —, Tx_2, Obs Group 4: —, —, Obs	Requires a larger sample size than two-group studies; random assignment to treatments is essential.
16. Combined experimental and ex post facto design	To study the possible effects of an experimenter-manipulated variable, a previously existing condition, and the interaction between the two	Group Time → Group 1: Exp_a, Random Assignment Group 1a: Tx_a, Obs Group 1b: Tx_b, Obs Group 2: Exp_b, Random Assignment Group 2a: Tx_a, Obs Group 2b: Tx_b, Obs	Requires a larger sample size than two-group studies; random assignment to the experimenter-manipulated variable is essential.

TABLE 9.1 Summary of Experimental and Ex Post Facto Designs

		Graphic Depiction	Comments on the Design
		Pre-Experimental Designs	
1. One-shot experimental case study	To show that one event (a treatment) precedes another event (the observation)	Group / Time → Group 1 | Tx | Obs	Shows a before-and-after sequence but cannot substantiate that this is a cause-and-effect relationship.
2. One group pretest–posttest design	To show that change occurs after a treatment	Group / Time → Group 1 | Obs | Tx | Obs	Provides a measure of change but yields no conclusive results about the cause of the change.
3. Static group comparison	To show that a group receiving a treatment behaves differently than one receiving no treatment	Group / Time → Group 1 | Tx | Obs Group 2 | — | Obs	Fails to determine pretreatment equivalence of groups.
		True Experimental Designs	
4. Pretest–posttest control group design	To show that change occurs following, but only following, a particular treatment	Group / Time → Random Assignment: Group 1 | Obs | Tx | Obs Group 2 | Obs | — | Obs	Controls for many potential threats to internal validity.
5. Solomon four-group design	To investigate the possible effect of pretesting	Group / Time → Random Assignment: Group 1 | Obs | Tx | Obs Group 2 | Obs | — | Obs Group 3 | — | Tx | Obs Group 4 | — | — | Obs	Enables the researcher to determine how pretesting may affect the final outcome observed.
6. Posttest-only control group design	To determine the effects of a treatment when pretesting cannot or should not occur	Group / Time → Random Assignment: Group 1 | Tx | Obs Group 2 | — | Obs	Uses the last two groups in the Solomon four-group design; random assignment to groups is critical for maximizing group equivalence.
7. Within-subjects design	To compare the relative effects of different treatments for the same participants	Group / Time → Group 1 | Tx_a | Obs_a | Tx_b | Obs_b	Useful only when effects of each treatment are temporary and localized.
		Quasi-Experimental Designs	
8. Nonrandomized control group pretest–posttest design	To show that two groups are equivalent with respect to the dependent variable prior to the treatment, thus eliminating initial group differences as an explanation for posttreatment differences	Group / Time → Group 1 | Obs | Tx | Obs Group 2 | Obs | — | Obs	Differs from experimental designs because test and control groups are not totally equivalent; equivalence on the pretest ensures equivalence only for variables that have specifically been measured.

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?

_____

_____ 4. If the researcher is *not* manipulating one or more independent variables, what precautions is the researcher taking to control for confounding variables? For example, is the researcher:

- Using one or more pretests to assess before-treatment group equivalence? If so, what variables are being pretested?

_____

_____

- Identifying matched pairs? If so, on the basis of what variables?

_____

_____

- Statistically controlling for confounding variables? If so, which ones?

_____

_____ 5. If the researcher is conducting a single-group or single-subject study, is the researcher:
- Conducting a series of observations both before and after the intervention (a time-series design)? _____
- Alternating either between two or more treatments or between treatment and nontreatment, with a new observation being made after each treatment or nontreatment (a reversal design)? _____
- Beginning an intervention at different times for different individuals or different contexts (a multiple baseline design)? _____

_____ 6. What other variables in the study (either identified or not identified by the researcher) might potentially affect the dependent variable? _____

_____

_____

_____ 7. To what extent might each of the following factors threaten the study's internal validity? If any of these factors pose a potential threat, how is the researcher minimizing or eliminating its influence? (Refer to Figure 9.1.)

History:

_____

Maturation:

_____

Testing:

_____

Instrumentation:

_____

Statistical regression:

_____

Selection:

_____

Attrition:

_____

_____ 8. With your answers to the preceding questions in mind, explain whether the study's results justifiably demonstrate a cause-and-effect relationship:

_____

_____

_____

_____

Meta-Analyses

Remember, we can conclude that a cause-and-effect relationship exists between an independent variable and a dependent variable only when 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:* A research study should be repeatable. In fact, 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 yield consistent, predictable results. A meta-analysis is primarily a statistical technique, and so we describe this procedure in greater depth in Chapter 11.

Conducting Experiments on the Internet

USING TECHNOLOGY

In Chapter 8 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 names 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. 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 three 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 (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.[4] (In case you're curious, Norwegians are especially partial to spicy Mexican food.)

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 8 for online questionnaires are equally applicable to online experiments (see the Practical Application "Using the Internet to Collect Data for a Descriptive Study" on pp. 205–206).

[3]As noted in Chapter 8, you can reach the site by going to the website of the Association for Psychological Science (www.psychologicalscience.org); click on "Psychology links," scroll down to "Other Sites of Interest," and then click on "Online Psychology Experiments." Alternatively, you can go directly to the site, which, as this book goes to press, is located at psych.hanover.edu/research/exponnet.html.

[4]As you might guess, the attrition (dropout) rate was higher for participants with longer between-session delays. To determine the extent to which the differing attrition rates for different treatment groups might jeopardize the study's internal validity, the researchers collected basic demographic data at the beginning of Session 1. During the data analysis, the researchers found no significant differences in any demographic variables or in learning speed between participants who completed all three sessions and those who did not.

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 discuss these procedures in Chapter 11).

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.

CONCEPTUAL ANALYSIS EXERCISE Identifying 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 suggested readings 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).

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. The researcher 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 her own words, the researcher's problem statement was 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 is not one of the designs we have specifically described in this chapter. Instead, it involves 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	Tx1	Obs	Obs
	Group 2	Tx2	Obs	Obs
	Group 3	Tx3	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.

Example 6

METHODOLOGY

[After an introductory paragraph outlining the chapter's contents, the author describes the sample—students enrolled in maternity nursing courses at two universities—used in the study. Then, as she begins a discussion of her procedure, she explains that the experimental treatments were based on the Tennyson-Cocchiarella concept-teaching model (1986) and presents the key elements of the model. We pick up the methodology chapter at the point where the author describes the specific treatments used for each of the three treatment groups.]

Description of the Treatment Groups

[The author first explains that, for each of the three groups, treatment consisted of instruction in the basic concepts of fetal monitoring, plus additional instructional strategies, or "teaching variables," that differed for the groups.] . . . Starting with a basic class and adding new teaching variables to each treatment group, however, did require additional time. The length of time required for teaching the three treatment groups varied between 1 and 2 hours. These timeframes were established based on the results of the survey of baccalaureate nursing schools, in which 36% of the schools responding had less than 1 hour to teach fetal monitoring theory, and 52% had 1 to 2 hours (Kinnick, 1989).

The teaching variables for the first treatment group included labels and definitions, and presentation of best examples. According to Merrill and Tennyson, these variables usually include additional information needed to aid in the clarification and understanding of the concepts (Merrill & Tennyson, 1977, p. 100). Therefore, the design of this didactic presentation began with a very basic overview of physiology at the uterofetoplacental unit. Electronic fetal monitoring patterns are a reflection of uterofetoplacental physiology. Understanding the normal physiology and changes in the physiology that cause inadequate fetal oxygenation help the learner to identify the various patterns, and whether patterns are normal or abnormal. Understanding the physiology is also the basis to identifying appropriate nursing intervention which promotes normal physiology (reduction or even elimination of fetal distress) when abnormal patterns occur.

Comments

The author points out a possible confounding variable *in her study: The three forms of instruction took varying amounts of time.*

The survey to which the author refers was administered during a pilot study that she conducted prior to conducting the dissertation itself. She published the pilot study as a research article, which she cites here.

In this and subsequent paragraphs the author describes the treatment used for each treatment group; in a later "Procedure" section, she describes the general procedure she used to conduct the study. More often, a researcher will include a description of how each group was treated within the procedure section itself. Either approach is acceptable, however, as long as the writer makes the organization of the methodology section clear (e.g., through headings and subheadings).

When the classes were taught, the majority of students did not have any theory about the process of labor and delivery. In addition, they had not seen a fetal monitor. Methods of monitoring the fetus and a brief description and discussion of external versus internal monitoring, therefore, needed to be discussed. In addition, it was necessary to show the students a print-out of a fetal monitor as well as explain what the graphs meant. Before the basic concepts related to interpretation of the fetal heart could be taught, the student also needed to recognize critical characteristics of a contraction pattern as seen on a monitor strip. Contraction patterns can be a cause of physiological changes at the uterofetoplacental site. After these areas had been covered, the concept label, definitions, and best examples were presented . . . .

This description of what most students knew (and did not know) before instruction gives the reader greater confidence that the results observed after instruction (i.e., students' test performance) were probably due to the instructional treatments, rather than to any earlier learning experiences that the students may have had.

This 1 hour presentation included labels, definitions, best examples, and clarifying information. In the experience of this researcher, this presentation reflects closely the method for teaching fetal monitoring used in most schools of nursing, especially when the allocated time for teaching this content is limited. This treatment group is referred to as Group 1 throughout the study.

Notice that the author's notion of what is "traditional" instruction is based on her own experiences, and she says so here.

The second treatment group began with the same presentation used with the first treatment group, plus the addition of expository presentations for each major concept. An expository presentation was added after the labels, definition, and best examples of each set of coordinate concepts had been completed. For example, following the definition and display of the best examples of baseline fetal heart rate and its coordinate concepts, an expository presentation was done of the coordinate concepts. When that was completed, the concept of baseline variability was introduced and the same order of teaching variables was used. The addition of the expository presentations added approximately half an hour, so that this treatment group was scheduled for one and one-half hours. This group (labels, definitions, best examples and expository presentation) is referred to as Group 2.

After describing Group 1, the author proceeds to descriptions of Group 2 and then Group 3 in a logical and systematic fashion. The use of three subheadings (something along the lines of Treatment for Group 1 *or* Group 1 Instruction*) might have been helpful, however.*

By "expository presentation," the author means giving a short explanation or lecture about important ideas and concepts.

The design in Group 2 was chosen based on the results of Dunn's research (1984) on concept learning with college age students. . . . *[The author briefly describes Dunn's findings and their relevance for the instruction presented to Group 2.]*

A rationale for a particular experimental treatment strengthens any research report. A brief rationale can easily be incorporated into the description of procedures; a longer one should probably be presented earlier in the research report.

The treatment design for the third group used the same teaching variables as in Group 2, plus the addition of an interrogatory presentation to follow each expository presentation. This involved the addition of . . . transparencies specifically developed for the interrogatory presentation. When a fetal monitor pattern was shown on the screen, students were requested to compare it with their handout of definitions (list of critical characteristics) and best examples, and to identify the concept shown on the fetal monitoring pattern. This treatment design incorporated all of the teaching variables of the Tennyson-Cocchiarella concept-teaching model.

By "interrogatory presentation," the author means asking questions to assess students' understanding of, and ability to apply, what they have learned.

Development of the Instruments

[In this section, the author describes the tests she used to assess what participants knew about fetal monitoring following instruction, as well as a short questionnaire she used to determine the extent to which each participant knew something about fetal monitoring before instruction.]

Procedure

Prior to implementing this research, approval for the project was obtained from the Human Research Committee at the University of Colorado, and the Internal Review Board for Research at the University of Northern Colorado (Appendix E). The researcher then met with all students in each maternity nursing course during their first class to explain the research and ask their consent to participate. Consent forms were provided for each student (Appendix E). Once this process was completed, the research design was implemented.

Because the author conducted the study at two universities, she followed the necessary human research review procedures at both institutions.

Each maternity nursing course had three groups participating in the research. Students in each of the courses were randomly assigned to one of these three groups. One group received the instructional method described in the Tennyson-Cocchiarella model of concept attainment. A second group received the same instructional method with the exception of the interrogatory presentation. The third group had a didactic presentation using only labels, definitions, best examples and clarifying information. In other words, both the expository and interrogatory presentations were eliminated from the presentation for the third group. In both schools, the researcher taught all three methods. A script (or lecture) was developed for the researcher to use in all the treatment groups so that the content was the same in each group (Appendix F). The students were tested in a class session within 2 to 3 days following the class (treatment).

As noted earlier in Chapter 9, random assignment is one effective way of ruling out the possible effects of confounding variables.

The first group mentioned here ("one group") is actually Group 3, and the last ("the third group") is actually Group 1; this reversal might cause confusion for the reader.

The use of a "script" here would help the researcher teach the content similarly for all three treatment groups (except, of course, for the things she intentionally wanted to do differently for the three groups). Thus, it should help minimize any influences the researcher's hypotheses might have on her delivery of different instructional methods.

After the completion of the clinical experience of all groups in each university, a parallel form of the classification test was again administered. The sequence can be summarized as follows:

Class instruction → Posttest → Clinical Rotation → Delayed Test upon Completion of Clinical Rotation

This graphic display of the procedure used is a helpful summary for the reader.

In addition, each student was requested to keep a record of the number of contacts each of them had with fetal monitoring tracings, the context, and type of pattern (Appendix G). For example, the student may have been assigned to a labor patient who had a normal pattern. The contact, however, could have been in clinical conference where actual monitor strips of patients were discussed, or also in a prenatal clinic where a nonstress test was done on a patient. The purpose of keeping these records [was] to identify the number of interrogatory examples the students encountered clinically and the range of examples. This information [could] be compared with the post test results.

The author presumably asked students to keep such records as a way of helping her interpret any unexpected results related to the delayed (postclinical rotation) test. Keep in mind, however, that such self-reporting techniques, dependent as they are on participants' diligence and memories, will not always yield totally accurate information.

Ideally, none of the students were to have had any contact in the clinical setting before the instruction and first test were done. However, it was impossible to schedule all three treatments before students in each maternity nursing course were assigned to the clinical setting since they began their clinical experiences the second week of classes. A few students in this situation were assigned to patients with fetal monitors attached. Since they did not have any theory on fetal monitoring, they were not responsible for interpretation of fetal monitor patterns. However, staff nurses and/or clinical instructors may have demonstrated how to attach and detach the equipment and talked about tracings seen by each student on their individual patients.

Here the author points out a potential weakness in her study: Some students had additional exposure to fetal monitoring outside of the instruction she had given them in their respective treatment groups. The exposure was apparently minimal, however, and so probably did not jeopardize the quality of her study. Such honesty is essential *in any research report.*

Statistical Analysis

[The author continues with a discussion of the statistical analyses she used to compare the performance of the three groups.]

NOTE: Excerpt is from *Learning Fetal Monitoring Under Three Conditions of Concept Teaching* (pp. 58–69) by V. Kinnick, 1989, unpublished doctoral dissertation, University of Colorado, Boulder. Reprinted with permission.

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. (2008). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Boston: Allyn & Bacon.

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: Prentice Hall/Pearson Education.

Deming, S. N., & Morgan, S. L. (1993). *Experimental design* (2nd ed.). New York: Elsevier Science.

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.

Heath, D. (1995). *An introduction to experimental design and statistics for biology.* Bristol, PA: Taylor & Francis.

Kirk, R. E. (1994). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Montgomery, D. C. (2009). *Design and analysis of experiments* (7th 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.

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.

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists.* Cambridge, England: Cambridge University Press.

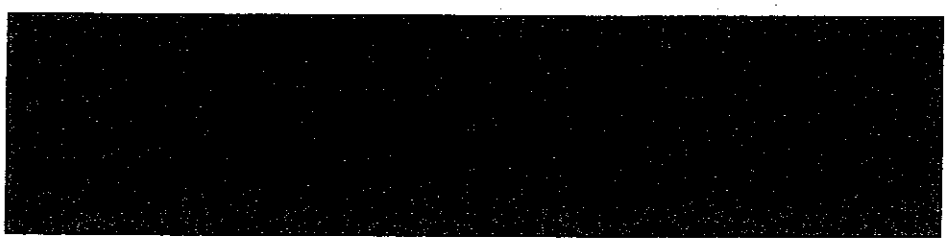
Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs.* Washington, DC: American Educational Research Association.

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.). Boston: Pearson/Allyn & Bacon.

ANSWERS TO THE CONCEPTUAL ANALYSIS EXERCISE "Identifying Research Designs" on page 251:

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 involving one 2-year-old girl is described. In this example, however, we see the approach being used with five horses, each of which is treated identically except for the date on which training begins.
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 the 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.