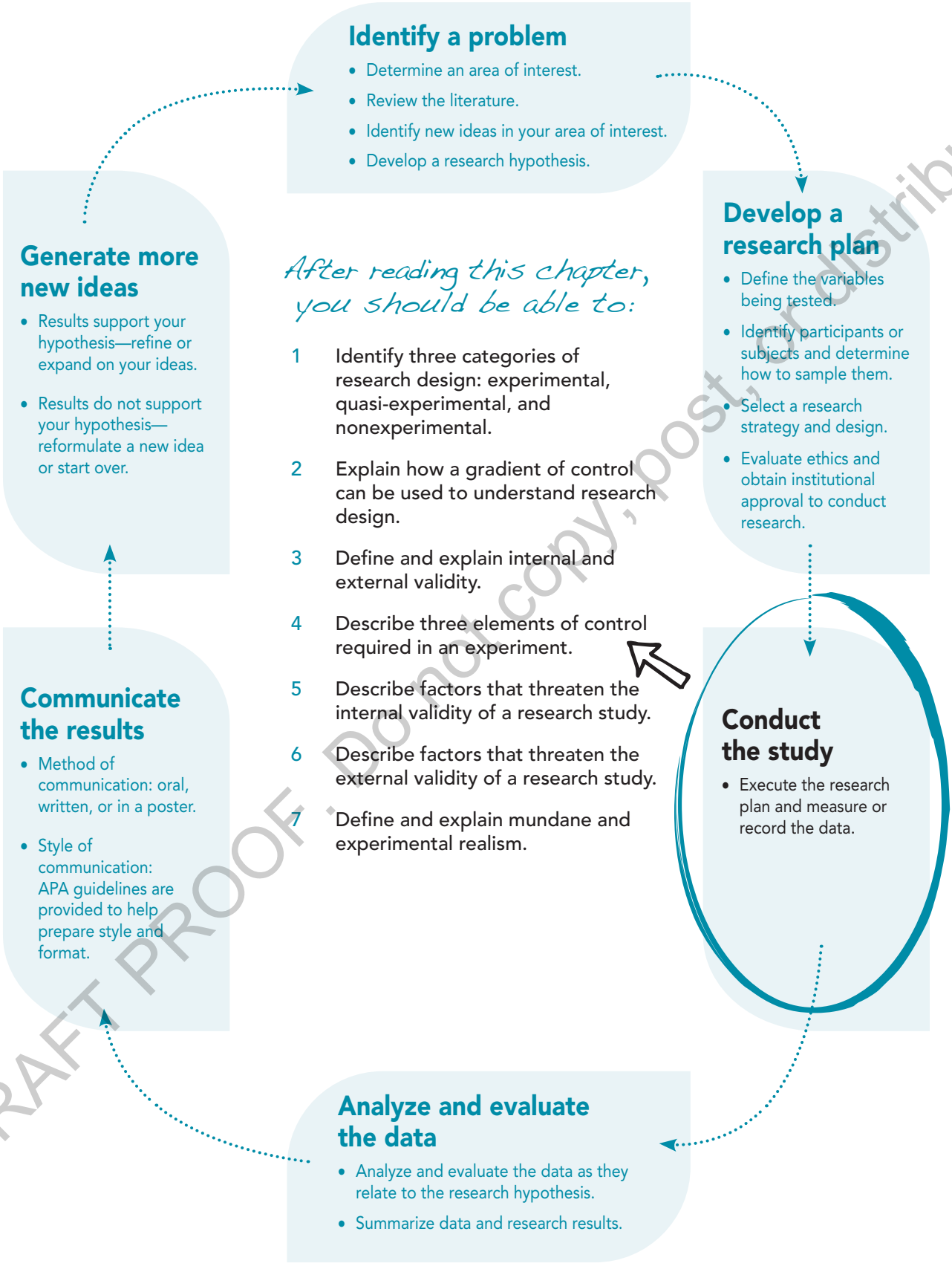


chapter
seven

CHOOSING A RESEARCH DESIGN

Education is about understanding the process of learning. You are in many ways an educational researcher when you ask questions about student learning and seek to answer those questions. For example, if you have students who are struggling in your class, you may ask, "How can I improve my teaching to increase my students' grades?" While you may not have considered the formal scientific process to answer your question, it can nonetheless be used to answer your question.

One way to think of research design is as a set of rules for how to make observations to answer questions. Each research design has a unique set of rules to help you control, manage, and organize what will be observed and how it will be observed. In this way, research design is similar to a board game, which has many rules to control, manage, and organize how you are allowed to move game pieces on a game board. Most board games, for example, have rules that tell you how many spaces you can move on the game board at most at a time and what to do if you pick up a certain card or land on a certain spot on the game board. The rules, in essence, define the game. Each board game only makes sense if players follow the rules.

Likewise, in science, the rules stated in a research design allow us to make sense of the conclusions we draw from the observations we make. In a board game, we follow rules to establish a winner; in science, we follow rules to establish conclusions from the observations we make. There are many ways in which you could observe your teaching—for example, videotape, observation reports by a teacher colleague or mentor, while using a specific teaching strategy, in a specific content area, or in a less controlled setting such as a library or study hall. In this chapter, we will explore the basic nature of the major categories of research design introduced in this book and organize how these research designs differ based on the types of conclusions they allow you to draw from the observations you make.

7.1 Designing a Study to Answer a Question

Conducting a study is important because it allows you to make observations using the scientific process to answer your research question. The type of study you conduct depends largely on the type of question you are asking. To conduct a research study, you need to be thoughtful of the extent to which you are actually answering your question. To illustrate, suppose you are a teacher and want to know if student collaboration promotes greater student learning. To answer this question, you organize one group of students into collaborative learning groups and a second group of students work as individuals. If you find that the students in the collaborative learning groups perform better, then can we conclude that the collaborative groups caused the difference between student learning outcomes?

The answer to your question depends on how thoughtful your research design was. For example, ability influences learning, so the student ability level in each group should be the same. Otherwise, if the ability level of one group of students was greater than the other, then maybe the ability level caused the differences between the two groups and not the collaboration. Other factors that could also influence learning as a result of collaboration include the content area, task being performed by the students, or gender. These additional factors should also be controlled to clearly show that the collaboration itself caused the differences between groups. In other words, designing a study is a careful, thoughtful, and often clever endeavor.

A **research design** is the specific methods and procedures used to answer a research question.

A research study applies specific methods and procedures, called the **research design,** to answer a research question. The types of research questions that you can ask are generally categorized as *exploratory*, *descriptive*, or *relational* questions. Each type of question is described with examples given in Table 7.1. In this chapter, we introduce many types of research designs used in the education sciences. In this book, Chapters 7 to 12 will describe in greater detail each type of research design introduced in this chapter.

LEARNING CHECK 1 ✓

1. State the type of question being asked for each example.
   - A. How often do students complete homework assignments?
   - B. What if the way that animals learn is similar to the way that humans learn?
   - C. Is family income related to enrollment in charter schools?

Answers: 1. A. descriptive, B. exploratory, C. relational.

Table 7.1 The Three Types of Questions That Researchers Ask

Type of Question	Question Stated	Description/Goal	Examples
Exploratory	"What if"	To "get an idea of" or "explore" an area of research that is not well understood. Rarely do these questions provide definitive answers; rather, they lead to a stronger focus for subsequent research.	1. What if a high-fat, high-sugar diet leads to attention-deficit hyperactivity disorder? 2. What if human memory has an infinite capacity for storage?
Descriptive	"What is" "How"	To characterize, explain, or "describe" variables that are related to a specific group of individuals. These questions are not concerned with relationships between variables; rather, they are concerned with simply describing variables.	1. What is the average amount of articulation errors in 4- to 5-year-old children? 2. How many minutes do elementary school-aged students spend watching TV per day?
Relational	"Does" "Is"	To determine the extent to which specified relationships exist between variables. These questions provide (1) causal explanations or (2) descriptions of the relationship between two or more variables.	1. Do low levels of serotonin in the brain cause depression? 2. Is free or reduced-price lunch status related to achievement?

7.2 Categories of Research Design

To answer a research question, you can choose a research design that falls into one of the following three categories, summarized in Figure 7.1:

- Experimental research design
- Quasi-experimental research design
- Nonexperimental research design

Each type of research design is distinguished by the level of control that is established in the design. The term **control** is used in research design to describe (a) the manipulation of a variable and (b) holding all other variables

Control in research design is (a) the manipulation of a variable and (b) holding all other variables constant. When control is low, neither criterion is met; when control is high, both criteria are met.

constant. When control is low, neither criterion (a) nor (b) is met. For example, suppose we observe play behavior among children at a park. The variable is play behavior; some children play quietly, and others play loudly. The children determine how loudly they play—the play behavior, then, is not manipulated or controlled by the researcher. Also, many other factors (e.g., the types of toys available to play with or the behavior of other children) can influence a child's play behavior at a park. Because the researcher does not manipulate the variable or hold these other variables constant, the study has low control.

Alternatively, when control is high, both criteria (a) and (b) are met. For example, to study play behavior, a researcher can have the children play one at a time on a playground. In one group, the children are told to play quietly; in another group, the children are told to play loudly. By manipulating the play behavior of the children (quiet play, loud play), the researcher establishes greater control. In addition, because all of the children play alone on the same playground, factors such as the types of toys available to play with or the behavior of other children are now held constant—all children, whether they play quietly or loudly, play alone with the same playground of toys. In this example, the researcher has established greater control by meeting both criteria (a) and (b) needed to establish control in a research design.

In this section, we introduce each research category, and we briefly describe the types of research designs that fall into each category. We will specifically distinguish between the levels of control established with each design because *control* is the key feature that can distinguish between categories of research design.

An **experimental research design** is the use of methods and procedures to make observations in which the researcher fully controls the conditions and experiences of participants by applying three required elements of control: randomization, manipulation, and comparison/control.

Experimental Research Designs

The staple of all designs is the experimental research design. The **experimental research design** is the use of methods and procedures to make observations in which the researcher fully controls the conditions and

Figure 7.1 The Three Categories of Research Design

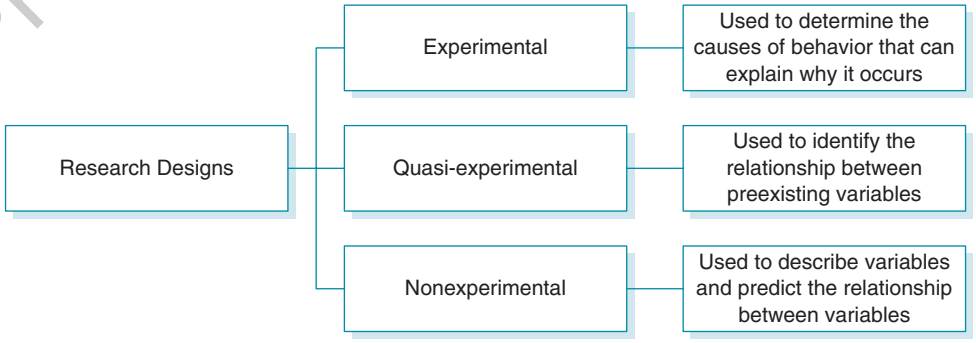
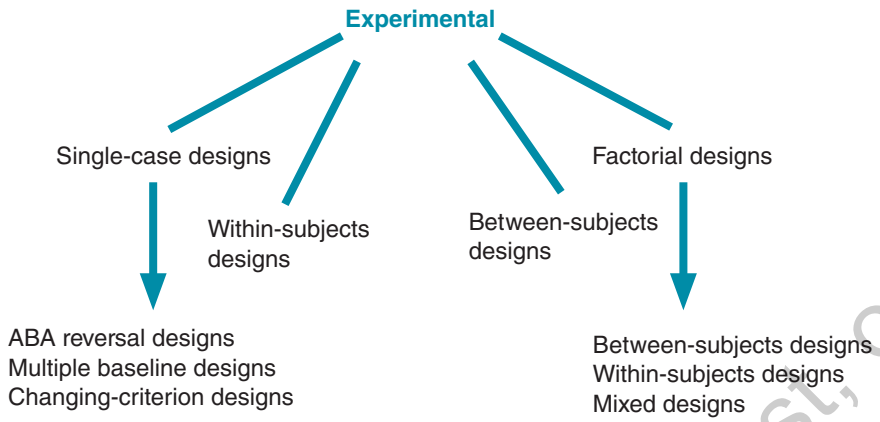


Figure 7.2 Experimental Research Designs



experiences of participants by applying three required elements of control: randomization, manipulation, and comparison/control. Each element of control is discussed in greater detail in Section 7.4. Figure 7.2 identifies many experimental research designs introduced in this book.

A key strength of the experimental research design is that it is the only research design capable of demonstrating cause and effect. To demonstrate that one factor causes changes in a dependent variable, the conditions and experiences of participants must be under the full control of the researcher. In the social sciences, such as psychology and sociology, this often means that an experiment is conducted in a laboratory and not in the environment where a behavior naturally operates. Suppose, for example, we study the effects of winning and losing on the desire to gamble. In a natural environment, it would be difficult to know if winning or losing causes changes in a person's desire to gamble because many other factors can vary in that setting. Some factors include the person's reasons for gambling that day, the amount of money available to gamble, the number of gamblers in a group, the types of games being played, and even the bright lights and sounds in the casino. We may be able to observe differences related to one's desire to gamble, but identifying the specific causes in that natural environment would be very difficult. To identify if winning or losing causes changes in a person's desire to gamble, Young, Wohl, Matheson, Baumann, and Anisman (2008) conducted an experiment by bringing students into a laboratory and having them experience a "virtual reality" casino in which the events, including winning and losing, were specifically controlled by the researchers. In this controlled setting of a virtual casino, the researchers specifically identified that high-risk gamblers have a much greater desire to gamble following a large win than following a series of small wins in a virtual casino setting. A key limitation of the experimental research design is that behavior that occurs under controlled conditions may not be the same as behavior that occurs in a natural environment.

Control is a key feature of research designs. Experimental designs have the greatest control over the conditions and experiences of participants.

In the education sciences, researchers rarely examine educational-related behaviors in a controlled lab setting. Our work primarily occurs in the natural educational setting—schools and classrooms. This can make experimental research more difficult in attempt to control and/or document those outside influences. For example, let's say you are a researcher exploring the difference between phonics-based and whole language–based reading instruction. You can control which student received what intervention and for how long at school, but you can't control parents from purchasing a "Getting Hooked on Phonics" CD from the Internet and using it with their child at home. Therefore, in educational research, researchers often include procedures to identify potential outside influences such as treatment diffusion or implementation fidelity.

Quasi-Experimental Research Designs

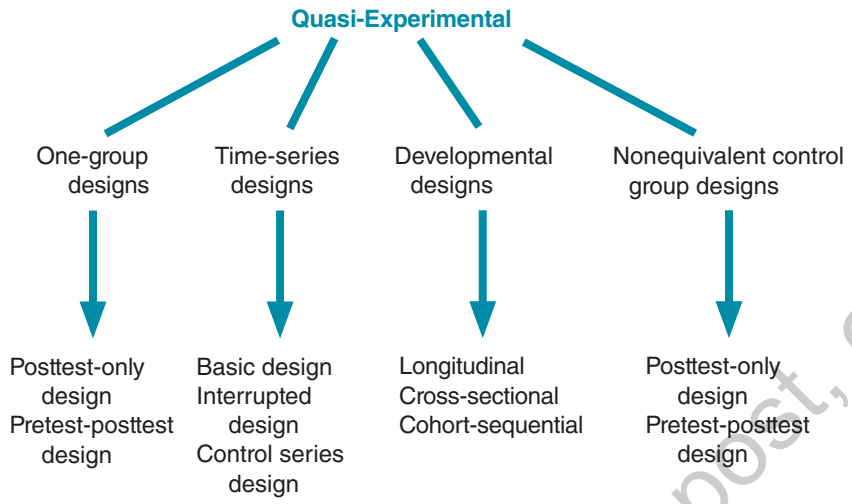
An alternative to the experimental research design for situations in which it is difficult or impossible to manipulate a factor is the quasi-experimental research design. The **quasi-experimental research design** is the use of methods and procedures to make observations in a study that is structured similar to an experiment, but the conditions and experiences of participants are not under the full control of the researcher. The conditions and experiences of participants are not under the full control of the researcher when the factor is not manipulated (i.e., it is "quasi-independent") or when the research design lacks a comparison/control group. This is probably the most common form of research in education because of the inability to randomly assign individual participants to treatment/comparison or control groups. Educational researchers instead assign intact classrooms or schools to treatment/comparison or control groups. Because of this lack of full control over the individual participants, this research is considered quasi-experimental. A **quasi-independent variable** is any factor in which the levels of that factor are preexisting. Quasi-independent variables of interest to researchers include characteristics of student participants, such as their gender (male, female), ability level (low, average, high), ethnicity, or free and reduced-price lunch status. When a factor is preexisting, participants cannot be randomly assigned to each level of that factor, but it allows researchers to study factors related to the unique characteristics of participants.

A **quasi-experimental research design** is the use of methods and procedures to make observations in a study that is structured similar to an experiment, but the conditions and experiences of participants lack some control because the study lacks random assignment, includes a preexisting factor (i.e., a variable that is not manipulated), or does not include a comparison/control group.

A **quasi-independent variable** is a variable with levels to which participants are not randomly assigned and that differentiates the groups or conditions being compared in a research study.

An example of an educational quasi-experiment is a study by Mann, Smith, and Kristjansson (2015), who investigated the effectiveness of the REAL Girls program on girls with preexisting problem behaviors that affected their school performance. Middle school girls from two public schools were recruited. The students from one middle school were assigned to participate in the REAL Girls program (treatment group) and students from the second middle school served as the control/comparison group and received delayed treatment. This study is considered quasi-experimental because the individual students were

Figure 7.3 Quasi-Experimental Research Designs



not randomly assigned to the groups. Assignment to the treatment group was by school, not by student. The researchers found that the REAL Girls program improved the academic self-efficacy, school connectedness, and identity of the girls.

A key limitation of the quasi-experimental research design is that researchers do not have full control to manipulate the independent variable (i.e., quasi-independent variable) and thus cannot demonstrate cause and effect. Referring back to the Mann et al. (2015) quasi-experiment, it is not possible to know if the REAL Girls program *caused* changes in academic self-efficacy, school connectedness, or identity because other factors, such as peer influences or home situations, could also be related to these factors and therefore could also be causing the changed attitude about school. Any time a factor is preexisting (i.e., quasi-independent), then any other factors related to it could also be causing changes in a dependent variable. Figure 7.3 identifies the many quasi-experimental research designs introduced in this book. Each research design is listed in Figure 7.3.

A **quasi-experimental research design** is structured similar to an experiment, but the conditions and experiences of participants are not under the full control of the researcher.

A **nonexperimental research design** is the use of methods and procedures to make observations in which the behavior or event is observed "as is" or without an intervention from the researcher.

Nonexperimental Research Designs

A common research design used in the education sciences is the nonexperimental research design. The **nonexperimental research design** is the use of methods and procedures to make observations in which the conditions or experiences of participants are not manipulated. A *manipulation* occurs when the researcher creates the conditions in which participants are observed; however, this is not always possible to study behavior. For example, we cannot manipulate the content of existing documents at different times in history, such as an analysis of Individualized Educational Plans (IEPs) under different special education laws.

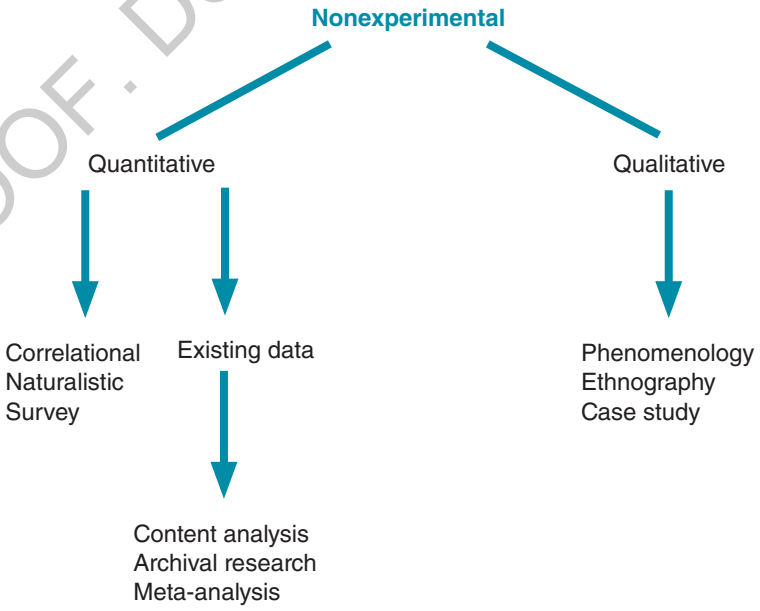
Another example is that we often cannot manipulate interactions in natural settings, such as those between a student and a teacher or between an athlete and a coach. Situations such as these are certainly worthy of scientific investigation, so nonexperimental research designs have been adapted to study these types of situations.

In many situations, we want to study behavior in settings where the behavior or variables being observed cannot be manipulated. Figure 7.4 identifies many nonexperimental research designs introduced in this book—each design will be introduced in greater detail in Chapters 8 and 9. A key characteristic that differentiates nonexperimental designs from all other research designs is that the behavior or event being observed is observed "as is" or without intervention from the researcher. For example, Durden, Escalante, and Blitch (2014) observed teachers in ethnically diverse preschool classrooms to examine how the teachers implemented culturally relevant pedagogical practices. In this study, teachers were observed as they went about their typical day for 1 year as the researchers observed the physical environment, teacher-child interactions, and nonverbal communications.

A key strength of the nonexperimental research design is that it can be used to make observations in settings in which the behaviors and events being observed naturally operate. Referring back to the Durden et al. (2014) study, these researchers observed 28 children and 51 teachers (graduate students and preservice teachers) in two preschool classrooms. Likewise, we can observe other situations in natural settings—that is, a prisoner and a guard in a prison or an athlete and a coach during a game. In each example, we make observations in a setting where the subjects or participants being observed would naturally interact.

A key limitation of the nonexperimental research design is that it lacks the control needed to demonstrate cause and effect. For example, if the classroom environment lacked

Figure 7.4 Nonexperimental Research Designs



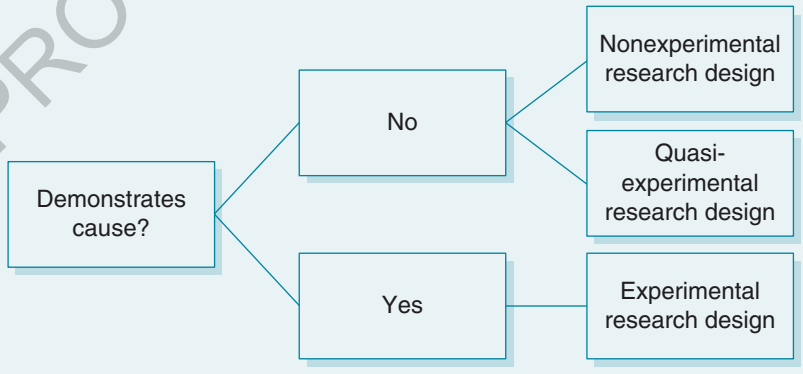
an ethnically and gender diverse library of books, toys, and games, we cannot know for sure that non–culturally relevant pedagogical practices caused that lack of diversity because other factors (e.g., monetary resources, classroom space, administrative decisions) could also explain the type of classroom supplies present. It is often difficult to anticipate all alternative explanations for what is observed in a natural setting. Using a nonexperimental research design, then, we can speculate about potential causes for the observations we make, but we cannot know for sure without greater control.

MAKING SENSE—"CAUSE" AS THE STANDARD OF RESEARCH DESIGN?

As shown in Figure 7.5, we could characterize a research design as either having demonstrated cause and effect (i.e., experimental research design) or having failed to establish the control needed to do so (i.e., nonexperimental and quasi-experimental designs). This characterization can lead to the erroneous conclusion that the best or superior research designs are those that demonstrate cause—this conclusion is not true.

Certainly, one of the goals in science discussed in Chapter 1 is to *explain* the causes of the behaviors and events we observe. An experimental research design is the only design capable of meeting this goal. However, it is also a goal in science to *describe* the behaviors and events we observe and to determine the extent to which we can *predict* their occurrence in different situations. Nonexperimental and quasi-experimental research designs are well adapted to meet these goals. Studying behavior is complex, and we must understand that not all behaviors and events can be brought under the full control of a researcher. Therefore, nonexperimental and quasi-experimental research designs are an essential and valuable tool that allows researchers to meet the goals of science and add to an understanding of the behaviors and events they observe.

Figure 7.5 Classifying Research Designs by Whether They Can Demonstrate Cause and Effect



7.3 Internal and External Validity

Categorizing research design is rather difficult. Indeed, there is not even full agreement among scientists about what types of research designs fit into each category. In other research methods textbooks, for example, many of the quasi-experimental designs listed in Figure 7.3 are instead taught as being nonexperimental designs. However, try not to get bogged down in the categorization of research design. Instead, use the three main categories of research design—experimental, quasi-experimental, and nonexperimental—as a way to organize the general types of designs used.

Internal validity is higher with greater control; external validity is higher with fewer constraints.

Categorization can oversimplify the complexity of research design. For example, you will find in Chapter 9 that the single-case design is taught as a type of experiment; however, not all researchers agree on this categorization. In Chapter 12, you will find that sometimes we can combine research designs that belong to different categories. The idea here is that thinking of research design only in terms of categories takes away from the true complexity of research design. A better approach is to think of research design along a gradient of control, as illustrated in Figure 7.6. Experimental research designs have the greatest control in that the conditions and experiences of participants are under the full control of the researcher. This control is less in a quasi-experimental research design and can be absent in a nonexperimental research design.

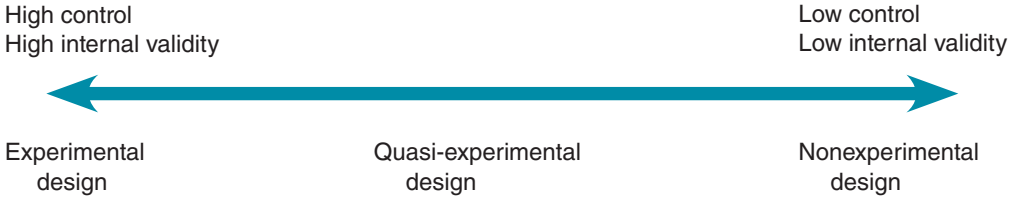
Internal validity is the extent to which a research design includes enough control of the conditions and experiences of participants that it can demonstrate a single unambiguous explanation for a manipulation—that is, cause and effect.

External validity is the extent to which observations made in a study generalize beyond the specific manipulations or constraints in the study.

The level of control in a research design directly relates to **internal validity** or the extent to which the research design can demonstrate cause and effect. The more control in a research design, the higher the internal validity. Experimental research designs have the greatest control and therefore the highest internal validity; nonexperimental research designs typically have the least control and therefore the lowest internal validity.

A second validity for research design, called **external validity,** relates to the generalizability of the study. Generalizability is an aspect of the research design that can constrain or limit observations to the specific conditions or manipulations in a study. An educational researcher, for example, may conduct a study in an

Figure 7.6 A Description of Research Design as a Gradient of Control



urban school setting with only males. The constraints in this study are the school setting and use of only male participants. Would the findings of this study in this situation be the same in other situations, such as in a rural setting or with females? The more an observation generalizes beyond the specific conditions or constraints in a study, the higher the external validity. The fewer the constraints or the more natural the settings within which observations are made, the higher the external validity of a research study tends to be.

LEARNING CHECK 2 ✓

1. State the three categories or types of research design.
2. State the category of research design that can demonstrate cause and effect.
3. Which type of research design has the highest internal validity?

Answers: 1. Experimental, quasi-experimental, and nonexperimental research design; 2. Experimental research design; 3. Experimental research design.

7.4 Demonstrating Cause in an Experiment

Any study that demonstrates cause is called an **experiment.** To demonstrate cause, an experiment must follow strict procedures to ensure that all other possible causes have been minimized or eliminated. Therefore, researchers must control the conditions under which observations are made to isolate cause-and-effect relationships between variables. Figure 7.7 uses an example to show the steps of a typical experiment. We will work through this example to describe the basic structure of an experiment.

An **experiment** is the methods and procedures used in an experimental research design to specifically control the conditions under which observations are made to isolate cause-and-effect relationships between variables.

Three required elements of control in an experiment are randomization, manipulation, and comparison/control. Each element of control is described further in this section.

General Elements and Structure of Experiments

An experiment includes three key elements of control that allow researchers to draw cause-and-effect conclusions:

1. Randomization (random sampling and random assignment)
2. Manipulation (of variables that operate in an experiment)
3. Comparison/control (a control group)

The hypothetical experiment illustrated in Figure 7.7 illustrates a hypothetical experiment to determine the effect of distraction on student test scores. To employ **randomization,** we use

Randomization is the use of methods for selecting individuals to participate in a study and assigning them to groups such that each individual has an equal chance of being selected to participate and assigned to a group.

Researchers use randomization to ensure that individuals are selected to participate at random (random sampling or random selection) and are assigned to groups at random (**random assignment**).

An **independent variable** or **factor** is the variable that is manipulated in an experiment. The levels of the variable remain unchanged (or "independent") between groups in an experiment. It is the "presumed cause."

The **levels of a factor** are the specific conditions or groups created by manipulating that factor.

In an experiment, random assignment is used to control for individual differences.

Individual differences are the unique characteristics of participants in a sample that can differ from one participant to another.

A **confound** or confounding variable is a variable not accounted for in a research study that could be causing or associated with observed changes in the independent variable(s).

The **dependent variable** is the variable that is believed to change in the presence of the independent variable. It is the "presumed effect."

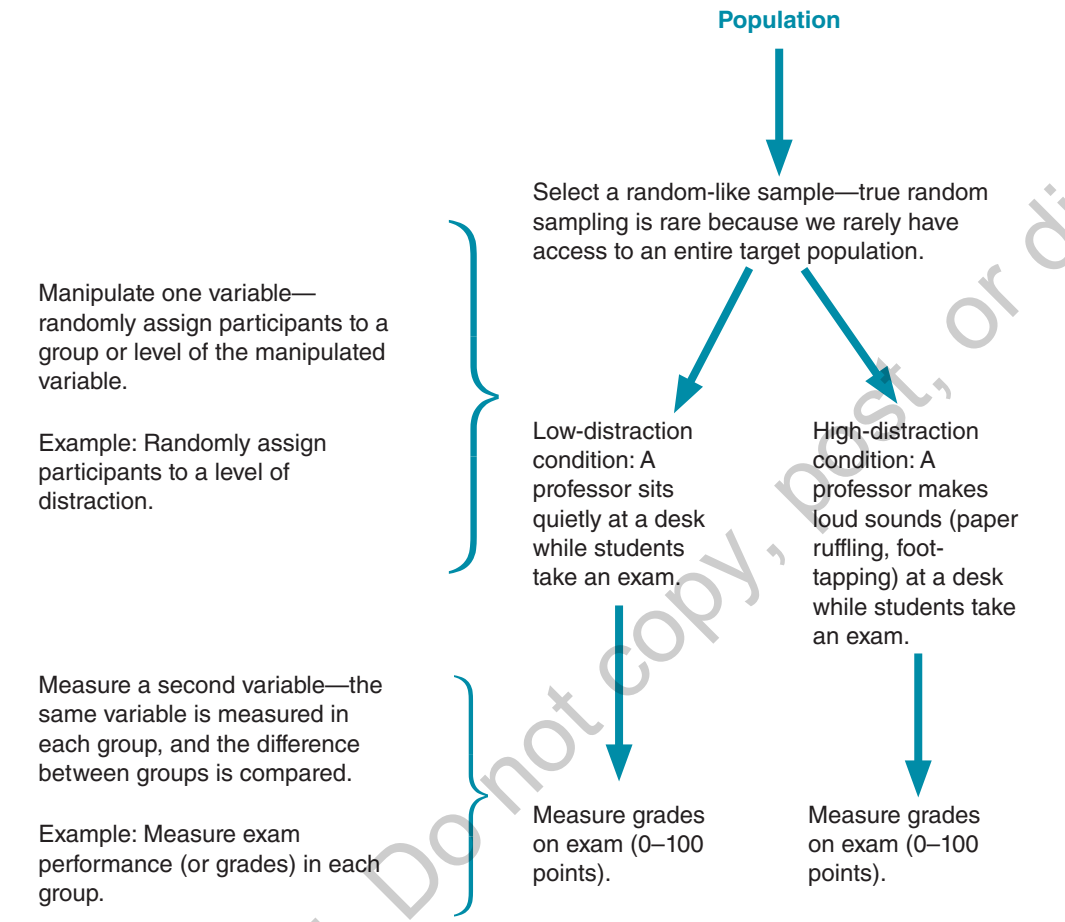
random sampling by selecting a sample at random from a population of students, and we then use **random assignment** to assign students to one of two groups at random. In one group, the teacher sits quietly while students take an exam (low-distraction condition); in the other group, the teacher rattles papers, taps her foot, and makes other sounds during an exam (high-distraction condition).

Random sampling is a method of selecting participants such that all individuals have an equal chance of being selected to participate. Random assignment is a method of assigning participants to groups such that each participant has an equal chance of being assigned to each group. To use random assignment, we identify the **independent variable** or **factor** that will be manipulated in an experiment (note that *manipulation* is the second element of control in an experiment). We then assign participants to each **level** of that factor using a random procedure, such as using a random numbers table to assign participants to groups. (Note that a random numbers table is given in Appendix B.1 with instructions for how to use it.) As shown in Figure 7.7, in our experiment, we manipulated "distraction" (the factor), which has two levels (low, high). We then randomly assigned participants to one level or the other. Each level of the independent variable is a group in our design.

Random assignment was first introduced in research with plant seeds (Fisher, 1925, 1935) and has since been applied to research with humans. What was learned in studies with plants is that random assignment controls for the **individual differences** in the characteristics of plants, and the same principle can be applied to human participants. An individual difference is any characteristic that can differentiate people, including their ethnicity, gender, ability level, free and reduced-price lunch status, age or grade, school attendance, or any other characteristic that may differ between people in a specific study.

We use random assignment with humans to control for individual differences in participant characteristics by ensuring that the characteristics of participants in each group of an experiment vary entirely by chance. If we do not control for individual differences, then any number of participant characteristics could differ between groups and explain an observed difference between groups. The individual differences would be a **confound,** or an alternative explanation for an observation in an experiment by affecting the independent variable. In our hypothetical experiment regarding distraction during an exam, a possible confounding variable may be hearing acuity. If students from one group had lower hearing acuity than the other, it may serve as an alternate explanation of the independent variable.

Figure 7.7 The Basic Structure of a Hypothetical Experiment That Includes Randomization, Manipulation, and Comparison



We create at least two groups in an experiment so that a presumed cause (high distraction) can be compared to a group where it is absent or minimal (low distraction). We can then compare grades in each group to determine the difference or effect that distraction had on exam grades. The measured variable in an experiment is called the **dependent variable.** If a difference is observed between the low- and high-distraction groups, then we conclude that distraction levels caused the difference because we used randomization, manipulation, and comparison/control to design the experiment.

An extraneous variable is one that is not controlled for in the study that affects the dependent variable.

Uncontrolled variables that directly affect the dependent variable are called extraneous variables. Like confounding variables, they add error to the experiment by affecting the results of the study. Possible extraneous variables in our hypothetical study of distraction on test scores may be tiredness, time of day, or student interest in the content area. Additional factors to be considered to draw cause-and-effect conclusions are described in Section 7.6.

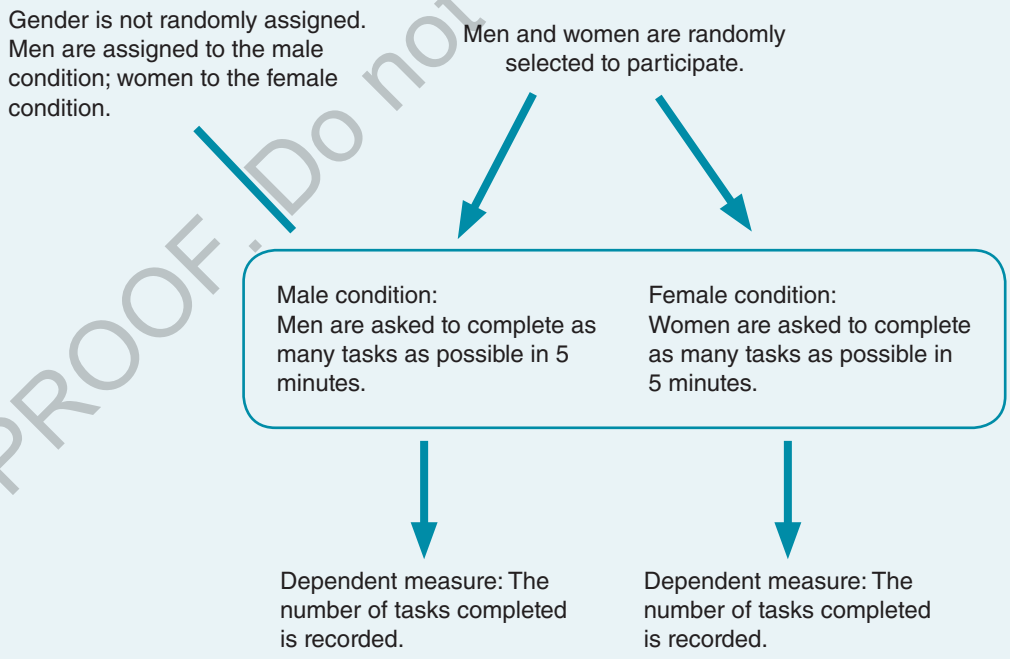
In an experiment, the independent variable is manipulated to create groups; the dependent variable is measured in each group.

MAKING SENSE—DISTINGUISHING BETWEEN AN EXPERIMENT AND A QUASI-EXPERIMENT

A researcher must manipulate the independent variable or factor in an experiment. Manipulating the factor means that the researcher creates the levels of that factor so that participants can then be assigned to a level or group at random. If the researcher does not manipulate the levels of the factor, then participants cannot be randomly assigned to groups, and the study is not an experiment. When a factor is not manipulated, the factor is called a quasi-independent variable.

Quasi-independent variables can be readily identified because these factors are typically characteristics that are unique to participants. For example, suppose we measure differences in the number of tasks completed by men and women. Figure 7.8 illustrates this study, which at first glance appears to be an experiment. However, gender is a characteristic of the participants and cannot be randomly assigned, which makes this factor a quasi-independent variable and this study a quasi-experimental research design. Be careful, therefore, to identify when the levels of a factor are manipulated because this one change can influence whether a study is experimental (demonstrates cause and effect) or quasi-experimental (does not demonstrate cause and effect).

Figure 7.8 A Quasi-Experimental Research Design



Note: Although experimental procedures are used, the factor (gender) was preexisting, which makes this a quasi-experimental research design.

7.5 Ethics in Focus: Beneficence and Random Assignment

In an experiment, researchers manipulate the levels of an independent variable and randomly assign participants to groups to establish control. Researchers also include a control or comparison group. In the example illustrated in Figure 7.7, a high-distraction condition was compared to a low-distraction condition. This example is likely associated with little ethical concern because manipulating the levels of distraction did not necessarily result in significant benefits or risks to participants. However, some situations may produce big differences in how participants are treated in each group. In these situations, there can be an ethical concern that relates to *beneficence*, which is the equal distribution of potential costs and benefits of participation (see Chapter 3).

Random assignment ensures that all participants in the research study have an equal chance of being assigned to a group and, therefore, an equal chance of receiving whatever benefits and costs are associated with participation in that group. Random assignment, however, may not be sufficient when one group has obviously greater benefits than another group. For example, studies that examine effectiveness of different instructional methods can have significant benefits for those participants receiving a superior type of instruction; thus, the control group (no instruction condition) or a comparison group (alternate or typical instruction) can be viewed as relatively disadvantaged. In these situations, researchers will often compensate the disadvantaged group, such as giving the control or comparison group access to the superior instruction at some time after the study, referred to as *compensatory equalization of treatments* (see Kline, 2008) or delayed treatment. Such compensation is provided to participants to meet the ethical standard of beneficence, as required in the American Educational Research Association (2011) code of conduct.

Random assignment ensures that participants have an equal chance of receiving the benefits or taking the risks associated with participation in a group.

LEARNING CHECK 3 ✓

1. State three elements of control in an experiment that allow researchers to draw cause-and-effect conclusions.
2. An educational researcher tests whether attitudes toward full inclusion of students with disabilities in general education classes differ based on grade band (elementary, middle, or high school). Identify the independent variable and the dependent variable in this example.
3. A researcher examining the effects of a new math intervention fails to control for the reading ability of the students taking the math test. Is the reading ability of the students an extraneous or confounding variable?

Answers: 1. Randomization, manipulation, and comparison/control; 2. Independent variable: grade band. Dependent variable: Attitudes toward full inclusion; 3. Extraneous variable.

CONNECTING TO THE CLASSROOM

Conducting an experiment in an educational context is difficult because of the natural grouping of students into classrooms. In the context of a classroom setting, it is a challenge to assign individual students to receive different experiences. Here are some things to think about when thinking about undergoing a study within a classroom.

- Can the students be randomly assigned individually to a treatment or comparison group? The answer to this question is not always no, even within a classroom. For example, if you are interested in examining different ways to manage small group work, you could randomly assign individual students to a small group and then randomly assign the small groups different ways to operate (such as some groups with certain roles and rules and other groups without the roles and rules).
- Consider how much control you can have over the intervention—who gets it and how much they get. An experiment about different types of homework would be more difficult to control than an experiment of different ways to manage small groups during a class activity. Homework is conducted outside of the teacher purview and therefore will have less control than a small group activity that occurs under the direction of the teacher.
- Recruit your students to be a part of your investigation. Prior to beginning the study, inform the students what you plan to do and why. Avoid statements that predict that one form of intervention may be better than another; instead, simply want to know which intervention is better. Let them know that you will be seeking their input after the study. You can pass out a short survey or hold a focus group discussion to see what your students think.
- If conducting an experiment with treatment and comparison groups, make sure to treat each group equitably. Each group should receive an equal amount of your attention and feel important to the study. Both groups should get "something" (i.e., comparable treatment) rather than one group get something and the other group get nothing.
- If it is not possible to manage two different interventions in one classroom, then randomly assign different classrooms to the different interventions. Make sure the classrooms are "equal." You can do this if you teach the same course during two different class periods and students are essentially equal in the sense that they have the same type of students. For example, you will not be able to compare a class designated as AP (Advanced Placement) to a non-AP-designated class.

7.6 Threats to the Internal Validity of a Research Study

Validity was first introduced in Chapter 4 to describe measurement, or the extent to which a variable measures what it is intended to measure. In this chapter, we introduce validity to

describe research design, or the extent to which the claim of a researcher fits with what was actually observed in a research study. Factors that threaten (i.e., decrease) the internal validity of a research study are those factors that vary systematically with an independent variable. (Internal validity was introduced in Section 6.3 in this chapter.) Therefore, any threat to the internal validity of a study is a potential confound that must be controlled.

The following is a list of common threats to the internal validity of a research study, which are introduced in this section:

- History and maturation
- Regression and testing effects
- Instrumentation and measurement
- Heterogeneous attrition
- Environmental factors
- Treatment factors

Factors that threaten internal validity vary systematically with the levels of an independent variable.

History and Maturation

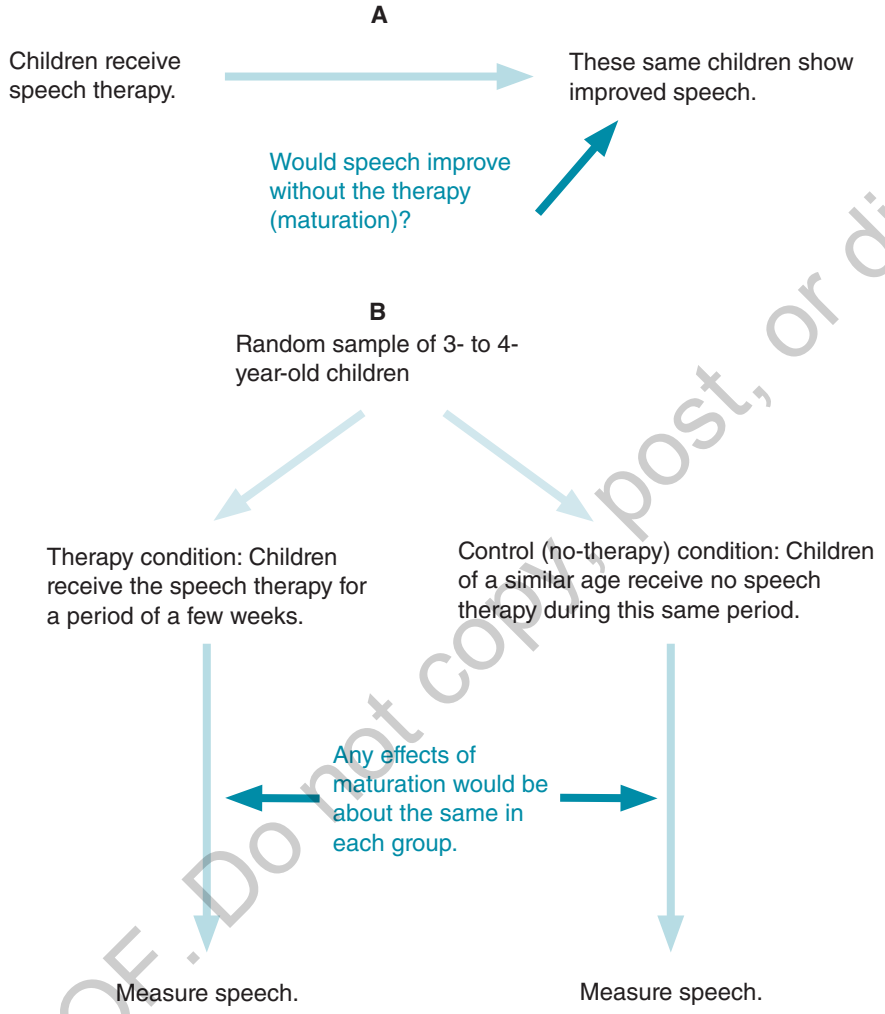
One threat to internal validity, called a **history effect**, refers to an unanticipated event that co-occurs with a treatment or manipulation in a study. History effects threaten internal validity when the event itself can also explain a research finding. For example, suppose researchers in New York City wanted to study the benefits of reading and so measured well-being on September 7, 2001, and again on September 14, 2001, among a group of participants who read each day for 1 hour during that time. A history effect, or unanticipated event, may have been the 9/11 terrorist attacks, which occurred during the study. If benefits of reading scores decreased, it was just as likely due to the 9/11 attacks (history effect) as it was to the reading manipulation. Other more subtle examples include holidays (e.g., measuring achievement after the winter break), school-related events (e.g., measuring instructional amount of instructional time after the Special Olympics or school pep rally), and public policy (e.g., measuring teacher stress levels before and after a change in education policy). In each case, 9/11, a holiday, school-related events, or public policy (history effects) can also explain any changes in benefits or reading, achievement, instructional time, or stress, respectively.

A **history effect** is a possible threat to internal validity in which an unanticipated event co-occurs with a treatment or manipulation in a study.

Maturation is a possible threat to internal validity in which a participant's physiological or psychological state changes over time during a study.

Another concern relates to **maturation**, which is a threat to internal validity in which a participant's physiological or psychological state changes over time during a study. Maturation refers to internal changes that exist within an individual and are not related to external events. Maturation includes factors such as age, learning, hunger, physical development, and boredom. As an example, suppose that a speech therapist shows that 3- to 4-year-old children improve their speech following her therapy, as illustrated in Figure 7.9. However, 3- to 4-year-old children develop speech naturally during that age period. Some

Figure 7.9 Maturation



(A) A child's speech could be due to maturation. (B) An experiment that controls for effects of maturation by including an appropriate control condition.

changes during the therapy, then, could simply be due to natural development and not to her specific therapy. One way to eliminate this problem would be to conduct an experiment that includes a no-therapy control condition, also illustrated in Figure 7.9.

Regression toward the mean is a change or shift in a participant's performance toward a level or score that is closer to or more typical of his or her true potential or mean ability on some measure, after previously scoring unusually high or low on the same measure.

Regression and Testing Effects

Some possible threats to internal validity are related to performance. Two examples of this are regression toward the mean and testing effects. **Regression toward the mean** occurs

when unusually high or low performance at one time shifts toward a level or score that is more typical or closer to the mean of an individual's true ability at a second time. You see this firsthand any time you obtain a better score on a makeup exam after "bombing" the first exam. One very possible explanation for the change in performance is regression toward the mean or toward one's true abilities.

Regression toward the mean usually occurs when participants are selected from the bottom or top percentile (extreme groups) in a population because initial scores will be unusually high or low for that group. For example, suppose you select a low-performing high school to examine changes in test scores of students after changing the start time to 1 hour later. You then give them a test before and after changing the school start time. As illustrated in Figure 7.10, without a control group that does not have the earlier start time, any improvement in scores on this test could be due to regression toward the mean.

A **testing effect** is the improved performance on a test or measure the second time it is taken due to the experience of taking the test.

A **testing effect** may be another explanation for the results. Testing effects occur when performance on a test or measure improves the second time it is taken. In the start time study, the improvement in test scores could be due to a testing effect inasmuch as participants may have learned something about the test the first time they took it. To distinguish between regression toward the mean and testing effects, keep in mind that regression toward the mean can be attributed to an increase or a decrease in performance from one time to another, whereas testing effects are attributed primarily to an increase in performance from one time to another. As illustrated in Figure 7.10, including an appropriate control group (i.e., a no-later start time group) can eliminate both threats to internal validity.

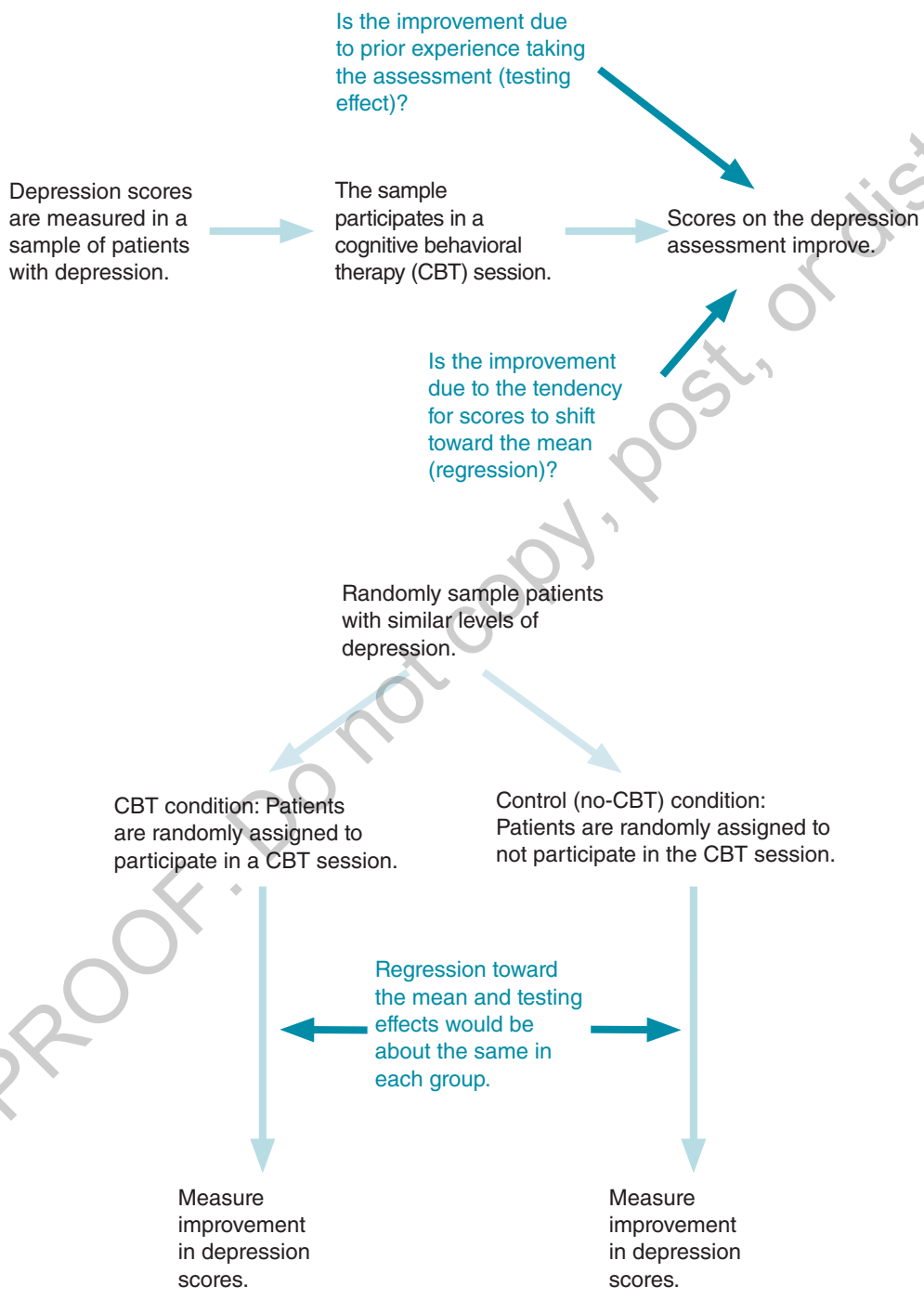
Instrumentation and Measurement

Sometimes an error in the measurement of a variable can threaten the internal validity of a research study. The possible threat of **instrumentation** refers to instances in which the measurement of the dependent variable changes due to low reliability of the instrument. Recall in Chapter 4 we discussed the reliability of measurements in terms of stability (test-retest), internal consistency (how items relate to each other), equivalence (having multiple forms of the same test), and interrater reliability (multiple raters). If the test does not measure the dependent variable consistently, whether it is over time such as a pre- and posttest or across different forms or raters, then the change in scores during a study may be due to the inadequate ability (i.e., low reliability) of the instrument and not due to a change in participants as a result of the independent variable.

Instrumentation is a possible threat to internal validity in which the measurement of the dependent variable has low reliability.

Instrumentation can be problematic when it is inherently prone to error, such as low reliability, but instrumentation may also be a problem due to inadequate training to use the measure. This can be a problem when the dependent measure requires some expertise to administer. For example, suppose three raters rate the time a participant held eye contact with a teacher during instruction. Learning to measure duration of eye contact should be a part

Figure 7.10 Regression Toward the Mean and Testing Effects



(A) Improved depression scores could be due to regression or testing effects. (B) An experiment that controls for the effects of both factors by including an appropriate control condition.

of the study protocol. If not properly trained, the raters will get better at rating the dependent variable (duration of eye contact) over time. One way to deal with getting better at administering the dependent variable is to intermix the order of the observations such that all of one condition is not run before another. If all female participants, for example, were observed prior to those in the male participants, then it is possible that ratings were better for the group that was observed last (the male group). In this case, instrumentation can threaten the internal validity of the study because the experience of the raters varies systematically with the levels of the factor, as illustrated in Figure 7.11.

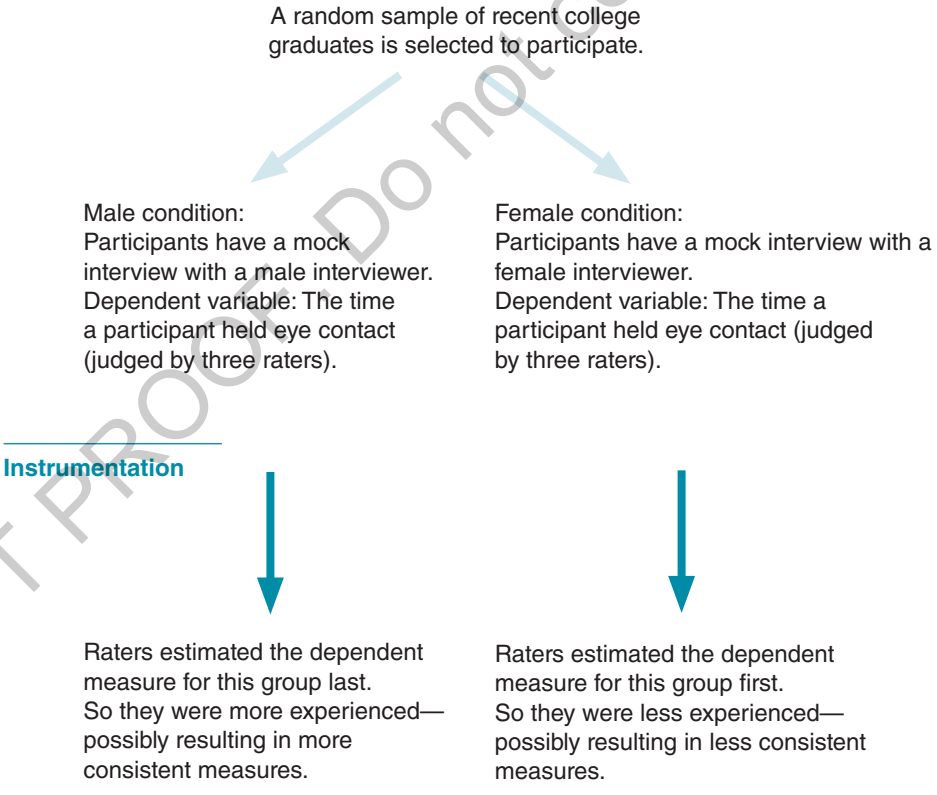
Attrition, or **experimental mortality**, is a possible threat to validity in which a participant does not show up for a study at a scheduled time or fails to complete the study.

Heterogeneous attrition is a possible threat to internal validity in which rates of attrition are different between groups in a study.

Attrition or Experimental Mortality

A common threat to internal validity can arise when a study is conducted across time such as multiple trials. days, or weeks. The problem of **attrition**, or **experimental mortality**, occurs when a participant does not show up for a study at a scheduled time

Figure 6.12 Instrumentation



The experience of the raters (instrumentation) varies systematically with the independent variable (gender of interviewer), which threatens the internal validity of the study.

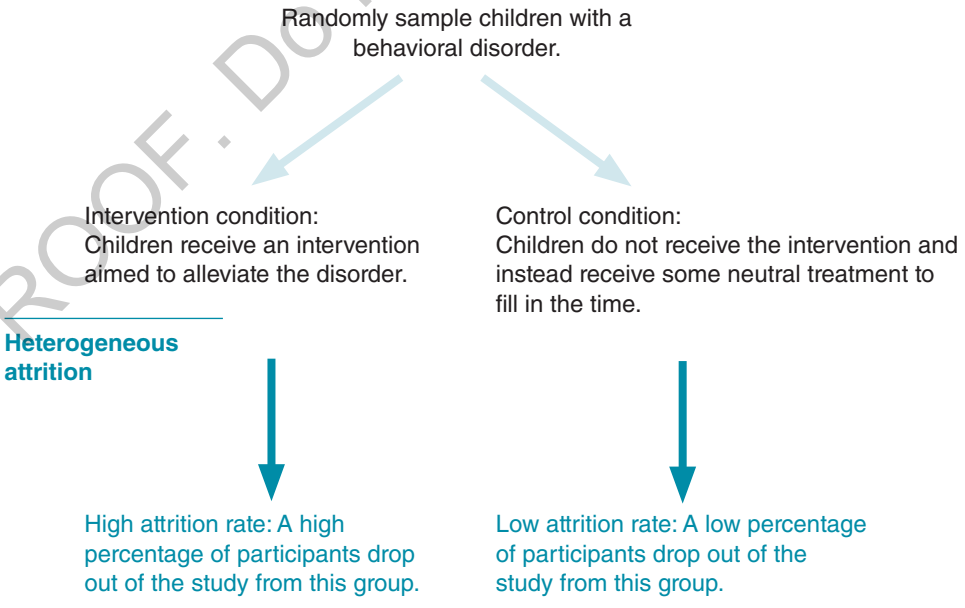
or fails to complete the study. A type of attrition that specifically threatens internal validity is called **heterogeneous attrition**, which occurs when attrition rates in one group are more or less than attrition rates in another group. Heterogeneous attrition is a threat to internal validity because attrition rates are different in each group. To illustrate, suppose you randomly assign children with a behavioral disorder to receive or not receive an intervention. In this case, the intervention group is likely to be a more tedious or demanding group for the children. If more children in the intervention than the no-intervention group drop out of the study, then attrition rates are now different between groups, and thus attrition rates vary systematically with the levels of the independent variable, as shown in Figure 7.12. Hence, the different rates of attrition can also potentially explain differences between groups in this example.

Another type of attrition, called **homogeneous attrition**, occurs when rates of attrition are the same in each group. Because attrition rates are the same in each group, homogeneous attrition does not threaten internal validity. It can, however, threaten the external validity of the study as defined and described further in Section 7.7.

Environmental Factors

Often, it is characteristics or dynamics of the study itself and the actions of the researcher that can a pose threats to internal validity. These types of threats are collectively referred to as *environmental factors*. Environmental factors include the time of day when a study is conducted, how researchers treat participants, and the location of the study.

Figure 7.12 Heterogeneous Attrition



Note: In this example, attrition rates vary systematically with the independent variable, which threatens the internal validity of the study.

Environmental factors can vary from one research design to another, so these factors should be carefully considered before conducting a study.

An environmental factor can only threaten the internal validity of a study when it varies systemically with the levels of an independent variable. Suppose, for example, you conduct a study to determine how participants judge the time they spend studying. In one group, participants rate their studying time each evening for a week; in a second group, participants rate their studying at the end of the week. In this case, if participants in one group had a test scheduled that week and those in a second group did not have a test scheduled that week, then this new factor (scheduled test) could threaten the internal validity of the study because it varies systematically with the levels of the independent variable (the rating format daily or weekly). Each environmental factor should be held constant so that only the rating format varies between groups.

Another way that the actions of the researcher can pose a threat to the internal validity of a study is when the researcher plays a role in the study itself. This is called experimenter effects. A researcher has a conflict of interest when it comes to the outcomes of the study. A researcher has spent time and effort in preparing for a study, and even if a hypothesis is not explicitly stated, there is an idea of what the outcome of the study may be. When the researcher then plays a role in the study such as conducting the educational intervention or collecting data on the dependent variable, this bias may alter the findings. For example, the researcher may alter effects of the intervention by being overly enthusiastic with the treatment group and not the control group or unconsciously help the students in the treatment group during the posttest. To combat this possibility, educational researchers should train others, called interventionists, to deliver the treatment and collect data. Many times, this could be the regular school teacher or a trained graduate student or other member of the research team.

Treatment Factors

Treatment factors can also affect the internal validity of a study when conducting an experimental study. These factors include threats called treatment replications, diffusion of treatment, and participant effects. The treatment replication threat to internal validity involves delivering the treatment for a long enough period of time after which the effect is expected to occur. If the treatment is not delivered long enough, the effect of the treatment may not be seen, even if the treatment is a successful one. For example, a researcher designs an intervention to use graphic organizers, understand social studies content, and deliver the graphic organizer treatment for only one social studies session. It is unlikely that the effect of the graphic organizer will be seen. The time it takes to see the desired effect varies with the intervention and the participants. Prior research conducted by other researchers and personal experience of the researcher can provide information on the length of time a specific treatment should be delivered.

A second treatment factor that can affect the interval validity of a study is called diffusion of treatment. This is when the treatment inadvertently gets delivered to the control group. Treatment diffusion may occur in two ways when participants are within close proximity to each other. One way that treatment diffusion can occur is when the person delivers the treatment in the view of all participants, both treatment and control. For example,

a researcher is delivering two different types of reading instruction to separate reading groups that occur within the classroom. While one group is receiving one type of reading instruction, the other group could be listening in or watching. A second way that treatment diffusion occurs is when the participants of different conditions (i.e., treatment and control groups) talk to each other about the study. Participants in the treatment group may share vital information about the treatment that changes the behavior of the participant in the control group. Another way that treatment diffusion can occur in research that is conducted in educational settings is when some form of the treatment is delivered to participants in the control group outside of the educational setting. Take the parent who purchased that "Getting Hooked on Phonics" CD for his or her child who happened to be in the whole language condition of a study. This would be treatment diffusion. Researchers should monitor for treatment diffusion whenever this possibility exists through observation and communication with the interventionists.

It is also possible for participants to alter their behavior as a result of simply being selected to participate in a research study. These effects are called participant effects or reactivity effects. Participant effects can be a potential threat to internal validity because the change in participant behavior is not due to the effect of the treatment but is related to just being a part of a research study. The most common of these effects are known as social desirability, the Hawthorne effect, John Henry (or compensatory rivalry) effect, and resentful demoralization. **Social desirability** is when participants change their behavior to put themselves in the best light. They respond in ways that are not true to themselves but in a way that they think would be more socially desirable. This is more likely to occur when the research involves less socially desirable topics such as bullying or academic cheating. The participants may respond in a more socially desirable way by not reporting that they were a perpetrator or victim of a bully or by minimizing the extent of cheating. The other forms of participant effects are related to which condition the participant is assigned. The **Hawthorne effect** occurs when the participants are selected as part of the treatment group and change their behavior because they believe they are getting special treatment. On the other hand, participants who are assigned to the control group may try harder because they were not selected to receive the special treatment. This is called the **John Henry** effect or compensatory rivalry. This is like the old Avis rental car commercial, where they say, "We are number two, so we try harder." Participants assigned to the control group may instead feel demoralized or resentful because they were not selected for the treatment group. This is called **resentful demoralization.** In this case, the participant's behavior changes because he or she is less motivated.

There are several ways to reduce the possibility of participant effects. One way is to less than fully disclose the object of the study so the participant does not know the exact behavior that is under study. We discussed this back in Chapter 3. The researcher will need to share enough about the study so the participant can weigh the risks and benefits of participating but may not know the exact behavior the researcher is interested in studying. Another way to reduce the possibility of participant effects is to offer an alternate or unrelated intervention so both groups get something. Table 7.2 summarizes the threats to internal validity that were described in this section and also describes the threats to external validity that will be introduced in Section 7.7.

Table 7.2 Internal and External Validity

Type of Validity	What Is Common Among Threats to This Validity?	What Are the Common Threats to This Validity?
Internal validity	All threats vary systematically with the levels of the factor or independent variable.	History, maturation, regression toward the mean, testing effects, instrumentation, heterogeneous attrition, environmental factors, and treatment factors that vary or are different between groups
External validity	All threats are held constant across groups in a study.	Sampling and participant characteristics, homogeneous attrition, research settings, timing of measurements, and the operationalization of constructs

LEARNING CHECK 4 ✓

1. What is characteristic of a factor that threatens the internal validity of a research study?
2. A teacher records scores for 10 students who took a midterm and a makeup midterm exam. She finds that scores improved on the makeup exam. Which two factors can likely threaten the internal validity of this result?
3. Explain why heterogeneous attrition, and not homogeneous attrition, is a threat to internal validity.
4. Name the different treatment factors that can affect internal validity.

Answers: 1. Factors that threaten the internal validity of a research study vary systematically with the levels of the independent variable; 2. Regression toward the mean and testing effects; 3. Heterogeneous attrition, but not homogeneous attrition, occurs when attrition rates differ between groups and, therefore, only heterogeneous attrition is a threat to internal validity because it varies systematically with the levels of an independent variable; 4. Treatment replications, diffusion of treatment, social desirability, Hawthorne effect, John Henry effect (compensatory rivalry), and resentful demoralization.

7.7 Threats to the External Validity of a Research Study

Threats to internal validity vary systematically with the levels of an independent variable. However, it can also be problematic when factors that are held constant between groups threaten the external validity of a study. Factors that threaten the external validity of a study limit the extent to which observations made by a researcher generalize beyond the constraints of the study. (External validity was introduced in Section 6.3 in this chapter.) Hence, the

factor that is held constant becomes the constraint to which observations are limited.

External validity is a broad term and can be subcategorized into at least four validities, each of which is described in Table 7.3. The following common threats to the external validity of a research study are described in this section:

Factors that threaten external validity are held constant across groups in a study.

- Population validity
- Ecological validity
- Temporal validity
- Outcome validity
- Treatment validity

Population Validity: Sampling and Participant Characteristics

Results observed in a study can sometimes be constrained to the sample. The extent to which results generalize to the population from which a sample was selected is called **population validity**. Researchers select samples to learn more about the populations from which the samples were selected. Sampling directly from the target population will result in the highest population validity. However, this sampling

Population validity is the extent to which results observed in a study will generalize to the population from which a sample was selected.

Table 7.3 Five Subcategories for External Validity

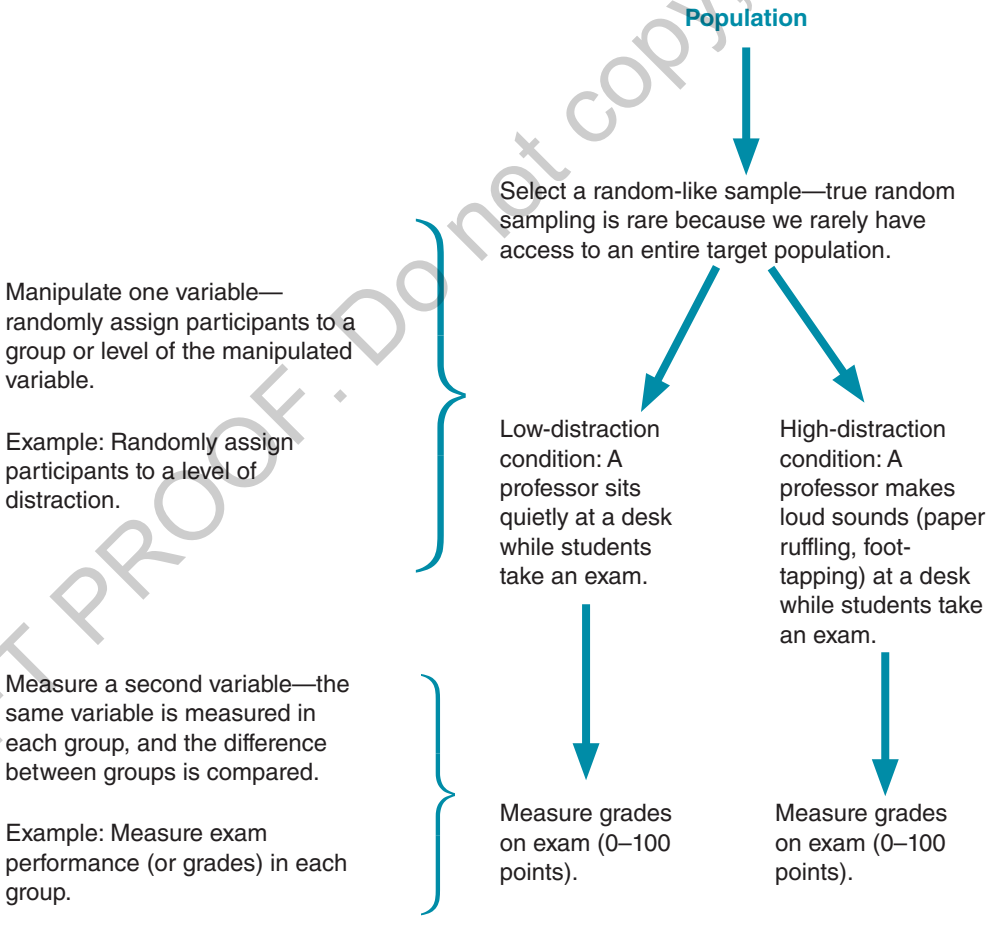
Subcategory of External Validity	Description	Threats to This Subcategory of External Validity
Population validity	The extent to which results observed in a study will generalize to the target population	Sampling methods and participant characteristics
Ecological validity	The extent to which results observed in a study will generalize across settings or environments	Research settings
Temporal validity	The extent to which results observed in a study will generalize across time and at different points in time	Timing of measurements, changes in our understanding of constructs over time
Outcome validity	The extent to which results observed in a study will generalize across different but related dependent variables	Operationalization of constructs
Treatment validity	The extent to which results observed in a study will generalize across different ways the intervention or treatment is conceptualized and administered	Implementing the treatment in different ways and varying amounts of time

method is often too difficult, so researchers more often select a sample of participants from a portion of the target population that is accessible, as described in Chapter 6, and illustrated here in Figure 7.13. When researchers select samples from an accessible population, they use strategies to ensure that characteristics in the sample are similar to those in the larger population, which will increase the population validity of a study.

Homogeneous attrition is a threat to population validity in which rates of attrition are about the same in each group.

One threat to population validity is **homogeneous attrition,** which occurs when the same number of participants from different conditions of a study does not show up at a scheduled time or fails to complete a study. In these cases, it is possible that participants who drop out or do not show up for a study are systematically different from those who do participate in the full study. Hence, the observations we make in the study will have low population validity in that results may be limited to only those participants who show up to

Figure 7.13



participate and may not generalize to those who do not. If differences between participants who complete and do not complete a study are related to changes in the dependent variable, this can lead to bias in the study (Goldkamp, 2008; Scott, Sonis, Creamer, & Dennis, 2006).

The key concern for population validity is that an effect that is observed in a study will only occur in that study. However, keep in mind that even when researchers use appropriate sampling methods, many results in a study can be constrained to a variety of factors even within a given population. For example, characteristics of teachers vary by grade band, and definitions of disabilities vary by state. If we study teacher characteristics or students with disabilities in the United States, then we must recognize that differences exist for these factors within the U.S. population. Issues of population validity, then, extend far beyond the methods used to select samples from populations. For this reason, it is important to be cautious in the extent to which we generalize observations to a larger population.

Ecological Validity: Research Settings

Results observed in a study can be constrained to the research setting in which observations were made. The extent to which results observed in a study will generalize across settings or environments is called **ecological validity**. For example, suppose a researcher has participants with autism listen to a story read in a monotone or dynamic voice in the quiet media center and finds that participants answer more literal recall questions when the words are spoken in a dynamic voice. Whether the results will generalize to other settings, such as in a self-contained classroom or general education classroom during the school day, determines the ecological validity of the research study.

Ecological validity is the extent to which results observed in a study will generalize across settings or environments.

Research conducted in a natural setting typically has high ecological validity because it is conducted in the same setting in which the behavior or event being measured would normally operate. In general, ecological validity is high so long as observations are not dependent on, or limited to, specific features of the research setting itself, such as the lab, the equipment used in the study, or the presence of the researcher.

Temporal Validity: Timing of Measurements

Results observed in a study can be constrained to the timing of observations made in a study. The extent to which results observed in a study will generalize across time and at different points in time is called **temporal validity**. The timing of measurements refers to the passage of time and to different points in time. The passage of time is illustrated by the phrase "Let me think about it." For example, students may change their mind about their choice of an academic major, or they may forget key information tested on an exam only to recall that information moments later. Temporal validity is the extent to which these observations (i.e., choice of future college major and recall on an exam) are stable, constant, or steady over time.

Temporal validity is the extent to which results observed in a study will generalize across time and at different points in time.

Temporal validity may also be related to how our explanations of behaviors change over time. As a result of research or changing attitudes, we may come to view or understand behaviors differently. We have many examples of this in education. One example of how ideas change over time is attention-deficit hyperactivity disorder (ADHD; Lange, Reichl, Lange, Tucha, & Tucha, 2010). In the 1960s, ADHD was called minimal brain dysfunction and referred to a class of learning problems presumed to be neurologically based and also included learning disability. This term was replaced in the 1970s with *hyperkinetic reaction in childhood* and recognized the hyperactivity aspect. In the late 1970s and early 1980s, the term *attention-deficit disorder* was introduced and distinguished with and without hyperactivity but was quickly replaced in the late 1980s with ADHD. This is the term we currently use to describe a persistent pattern of inattention and/or hyperactivity, but now there are three distinct subtypes: predominantly inattention, predominantly hyperactive, and combined. Autism and learning disability follow a similar pattern or change in understanding and diagnosis. We must consider these changes when evaluating the impact of temporal validity of research.

Outcome Validity: Operationalization of Constructs

Results observed in a study can be constrained to how the researcher defines the dependent variables. The extent to which results observed in a study generalize across related dependent measures for a variable or construct is called **outcome validity**. For example, if a study showed that a new behavioral intervention helped children stay on task, then it would have high outcome validity if it also showed that it reduced the number of times children disrupted the class. Disrupting class (outcome) is a different but related dependent variable to staying on task (outcome). As another example, if a study showed an effect of increased hunger, then it would have high outcome validity if it also showed an effect of increased calories consumed in a meal, for example. In this example, calories consumed in a meal (outcome) are a different but related dependent variable to hunger (outcome). Outcome validity, then, is the extent to which the outcomes or results of a research study can be generalized across different but related dependent variables. Hence, high outcome validity allows researchers to generalize a result or outcome beyond the specific measures used in a study.

Outcome validity is the extent to which the results or outcomes observed in a study will generalize across different but related dependent variables.

Treatment Validity: Operationalization of Treatments

Results of a study can be constrained by the extent to which a treatment is conceptualized and implemented. The extent to which the treatment can be generalized is called **treatment validity**. The question is whether the treatment can be implemented as it is conceptualized to other individuals. The key to treatment validity is the thorough description of the different components of the treatment, who

Treatment validity is the extent to which the results or outcomes observed in a study will generalize across different but related treatments.

delivered the treatment (such as teacher, paraeducator, or certified therapist), and dosage of the treatment (number and duration of treatments). Interventions that are difficult to implement or extend over a long period of time may be more difficult to generalize.

LEARNING CHECK 5 ✓

1. What is characteristic of a factor that threatens the external validity of a research study?
2. State five subcategories of external validity.
3. A researcher found that targeted professional development increased the job satisfaction of middle school teachers. Teachers' increased job satisfaction persisted for 6 months after the training. This study has high _____, which is a subcategory of external validity.

Answers: 1. Factors that threaten external validity are held constant across conditions in a study; 2. Population validity, ecological validity, temporal validity, outcome validity, and treatment validity; 3. temporal validity.

7.8 External Validity, Experimentation, and Realism

Researchers who conduct laboratory studies are aware that studies conducted in laboratories generally have low external validity, so they make efforts to increase the external validity of their studies. Because researchers can control all aspects of the study in the laboratory, laboratory studies tend to have high internal validity. To increase the external validity of laboratory studies, researchers can take additional steps to make the experimental situation *look* and *feel* as "real" as possible.

The extent to which an experimental situation *looks* real is called **mundane realism.** Suppose, for example, that you want to study gambling behavior. To establish mundane realism, you could create a casino-like setting in a laboratory with flashing lights, coin slots, and other games of risk. If the appearance of the setting looks real to participants, then the study has high mundane realism. Although field experiments, like one in an actual casino, will have higher mundane realism than laboratory experiments, efforts to mimic a "real" setting, such as a casino setting, can substantially increase the external validity of laboratory experiments.

Mundane realism is the extent to which a research setting physically resembles or looks like the natural or real-world environment being simulated.

Experimental realism is the extent to which the psychological aspects of a research setting are meaningful or feel real to participants.

The extent to which an experimental situation *feels* real is called **experimental realism.** In the casino gambling study, the more that participants feel as if they are in a casino during the study, the higher the experimental realism will be. If you set up a "real" casino-like setting in the laboratory, then it would likely have high mundane realism in that it *looks* like a real casino. However, if the study was conducted in a

laboratory or academic building, then participants may not entirely *feel* like they are at a real casino—because they realize where they are. In this way, it is important to reflect on both types of realism, as each type is distinct.

To enhance the experimental realism in a study, it is important that the manipulations in a study are meaningful to participants. For example, inherent physical abilities are meaningful to athletes, so we could manipulate high and low self-esteem by manipulating whether an athlete receives positive or negative feedback concerning his or her physical abilities. This manipulation would increase the experimental realism of the study because the manipulation is personally meaningful to participants. In all, making such considerations to increase the mundane realism, the experimental realism, or both in a study will increase the external validity of a research result.

Increasing the mundane and experimental realism of a study will increase the external validity of the study.

7.9 A Final Thought on Validity and Choosing a Research Design

Selecting a research design requires careful thoughtful planning, and some creativity. Be aware that few, if any, research designs will demonstrate high internal and high external validity in the same design. Indeed, some research designs have low internal validity, such as nonexperimental research designs, whereas others have low external validity, such as laboratory experiments. However, the goal in educational research is not to solve the world's educational problems in one study; this goal may not even be possible or realistic. Instead, the goal in educational research is to move forward and advance our knowledge of education and the behaviors and events that operate within it. Researchers are responsible for stating a question and choosing a research design that can answer their question. Researchers must choose an appropriate research design that can answer their question, and they must recognize the limitations of the research designs they choose.

Each research design used in behavioral research has strengths and limitations. Whether a study has high or low internal or external validity will vary from one study to another. For this reason, the greatest advancement of knowledge is found when many different types of research designs, with a complement of strengths and weaknesses, are employed to address the same problem. To advance knowledge, then, you do not have to design the perfect experiment; instead, you must choose an appropriate research design and be cautious to understand its strengths and weaknesses when drawing conclusions from the observations you make. In this way, to advance scientific knowledge, it is as important to be aware of the limitations and strengths of the research designs used to answer a research question.

High internal and external validity is not a prerequisite for "good" research designs. All research designs have limitations, and it is important that researchers recognize them.

Section III (Chapters 8 to 12) and Section IV (Chapters 13 to 15) will describe the research designs listed for each category of research design in Figures 7.2 to 7.4 in this chapter. You can revisit these figures as you read to help you organize how to think about research design in the chapters ahead.

CHAPTER SUMMARY

LO 1 Identify three categories of research design: experimental, quasi-experimental, and nonexperimental.

- A **research design** is the specific methods and procedures used to answer research questions. The types of research questions that researchers ask are generally categorized as exploratory, descriptive, or relational questions.
- A **nonexperimental research design** is the use of methods and procedures to make observations in which the behavior or event being observed is observed "as is" or without any intervention from the researcher.
- An **experimental research design** is the use of methods and procedures to make observations in which the researcher fully controls the conditions and experiences of participants by applying three required elements of control: randomization, manipulation, and comparison/control.
- A **quasi-experimental research design** is the use of methods and procedures to make observations in a study that is structured similar to an experiment, but the conditions and experiences of participants are not under the full control of the researcher. Specifically, the study includes a preexisting factor (i.e., a variable that is not manipulated: a **quasi-independent variable**) or lacks a comparison/control group.

LO 2 Explain how a gradient of control can be used to understand research design.

- Categorizing research can oversimplify the complexity of research design. Another way to approach research design is to think of it along a gradient of control. The more control present in a study, the more suited the design will be to demonstrate that one variable causes a change in a dependent variable. Studies with high control will be experimental; the less control in a study, the more quasi-experimental or nonexperimental the research design.

LO 3 Define and explain internal and external validity.

- **Internal validity** is the extent to which a research design includes enough control of the conditions and experiences of participants that it can demonstrate cause and effect.
- **External validity** is the extent to which observations made in a study generalize beyond the specific manipulations or constraints in the study.

LO 4 Describe three elements of control required in an experiment.

- An **experiment** has the following three elements of control that allow researchers to draw cause-and-effect conclusions:

  Randomization (random sampling and random assignment)

  Manipulation (of variables that operate in an experiment)

  Comparison/control (a control group)

- Randomization is used to ensure that individuals are selected to participate and assigned to groups in a study using a random procedure. Manipulation means that a researcher created the levels of the **independent variable,** thereby allowing the researcher to randomly assign participants to groups in the study. A comparison or control group is used to allow researchers to compare changes in a **dependent variable** in the presence and in the absence of a manipulation.

LO 5 Describe factors that threaten the internal validity of a research study.

- Factors that threaten the internal validity of a research study will vary systematically with the levels of an independent variable. These factors include **history effects, maturation, regression toward the mean, testing effects, instrumentation, heterogeneous attrition,** and environmental factors that can vary between groups in a study.

LO 6 Describe factors that threaten the external validity of a research study.

- Factors that threaten the external validity of a research study are those that are held constant across groups in a study. These factors include four subcategories of external validity:
  - **Population validity,** or the extent to which observations generalize beyond a sample to the population
  - **Ecological validity,** or the extent to which observations generalize across settings
  - **Temporal validity,** or the extent to which observations generalize across time or at different points in time
  - **Outcome validity,** or the extent to which observations generalize across different but related dependent variables

LO 7 Define and explain mundane and experimental realism.

- Mundane realism is the extent to which a research setting physically resembles or *looks* like the natural environment being simulated. Experimental realism is the extent to which the psychological aspects of a research setting are meaningful or *feel* real to participants. A study with high mundane and experimental realism will have high external validity.

KEY TERMS

attrition 203
confound 194
control 185
dependent variable 195
ecological validity 210
experiment 193
experimental mortality 203
experimental research design 186
external validity 192
factor 194
Hawthorne effect 206
heterogeneous attrition 203

history effect 199

homogeneous attrition 209

independent variable 194

instrumentation 201

instrumentation attrition 000

internal validity 192

John Henry (compensatory rivalry effect) 206

levels of a factor 000

maturation 199

nonexperimental research design 189

outcome validity 211

population validity 208

quasi-experimental research design 188

quasi-independent variable 188

random assignment 194

randomization 193

regression toward the mean 200

research design 184

resentful demoralization 206

social desirability 206

temporal validity 210

testing effect 201

treatment validity 211

REVIEW QUESTIONS

1. Choose the category of research design that best fits with the description given.
   A. Generally associated with high external validity
   B. Associated with the highest internal validity
   C. Structured as an experiment but lacks the control needed to demonstrate cause and effect
2. State the only category of research design that can demonstrate a cause-and-effect relationship between two factors.
3. In terms of controlling the conditions and experiences of participants:
   A. Which category of research design has the least control?
   B. Which has the most control?
   C. What is the relationship between control and internal validity?
4. State three elements of control that allow researchers to draw cause-and-effect conclusions.
5. Based on the following description of a hypothetical study, identify (a) the independent variable and (b) the dependent variable.

   A researcher believes that students will recall words that appear first and last on a list of vocabulary words more than the words that appear in the middle of the list. She presents each student with a list of 20 words to memorize in 3 minutes, and the words correctly recalled are recorded.
6. State whether each factor listed below is an example of an independent variable or a quasi-independent variable. Only state "quasi-independent variable" for participant variables that cannot be manipulated.

A. Marital status
B. Political affiliation
C. Amount of delay prior to recall
D. Type of school setting (urban, rural, suburban)
E. Time spent in reading instruction
F. Type of feedback (negative, positive)

7. What is characteristic of threats to internal validity? What is characteristic of threats to external validity?

8. A researcher measures the effectiveness of a drug intervention program by measuring the number of arrests of teenagers for drug possession before and after the program. One problem is that police initiate a crackdown on drugs in schools during this same time. What is the history effect in this example?

9. A researcher measures responsiveness to a drug treatment in high school students who volunteered or were mandated to participate. One problem that arises is that many students drop out of the program before the study is completed.
   A. What type of threat to validity does this example illustrate if dropout rates are the same among volunteer and mandated students? Is this a threat to internal or external validity?
   B. What type of threat to validity does this example illustrate if dropout rates differ between volunteer and mandated students? Is this a threat to internal or external validity?

10. Distinguish between regression toward the mean and testing effects as threats to internal validity.

11. Which subcategory of external validity is most likely threatened by homogeneous attrition? Explain.

12. A researcher uses an intervention program at a local youth center to help children with behavioral disorders. The researcher finds that the program was effective in an urban community but not in a rural community. What subcategory of external validity is low in this example? Explain.

13. A researcher measures a student's motivation to succeed as the amount of time spent studying. In a second study, the researcher conducts the same study but instead measures a student's motivation to succeed as the percentage of classes attended during a semester. Different results were observed in each study. What subcategory of external validity is low in this example? Explain.

14. State whether the following study has high mundane realism, high experimental realism, or both. Explain.

    A researcher measures gambling behavior among addicted gamblers. The study is conducted at a local casino (the researcher reserved a portion of the casino for

the duration of the study). She manipulated whether participants won or lost a predetermined game and recorded the amount of money participants gambled for 1 hour after this manipulation.

ACTIVITIES

1. A researcher hypothesizes that teachers will be more patient if they are also a parent. (a) Describe a research design to test this hypothesis. (b) Explain why you cannot choose an experimental research design for this example. Hint: Consider characteristics of quasi-independent variables.
2. Suppose you choose to conduct a study on fighting at school, eating behavior in school cafeterias, or safety concerns in high poverty area schools. Choose one topic and select and describe a research design.
3. Choose any research topic that interests you and state a research hypothesis. Identify the following information:
   A. Identify whether or not you will use an experimental research design to test your hypothesis. Explain.
   B. Identify factors that may threaten the internal validity of your study. Explain how your research design controls, or fails to control, for these threats to internal validity.
   C. Identify factors that may threaten the external validity of your study. Explain how your research design controls, or fails to control, for these threats to external validity.