PART VIII

# QUASI-EXPERIMENTAL RESEARCH

The prefix *quasi* means "resembling." Thus quasi-experimental research is research that resembles experimental research but is not true experimental research. Recall with a true between-groups experiment, random assignment to conditions is used to ensure the groups are equivalent and with a true within-subjects design counterbalancing is used to guard against order effects. Quasi-experiments are missing one of these safeguards. Although an independent variable is manipulated, either a control group is missing or participants are not randomly assigned to conditions (Cook & Campbell, 1979)[1].

Because the independent variable is manipulated before the dependent variable is measured, quasi-experimental research eliminates the directionality problem associated with non-experimental research. But because either counterbalancing techniques are not used or participants are not randomly assigned to conditions—making it likely that there are other differences between conditions—quasi-experimental research does not eliminate the problem of confounding variables. In terms of internal validity, therefore, quasi-experiments are generally somewhere between non-experimental studies and true experiments.

Quasi-experiments are most likely to be conducted in field settings in which random assignment is difficult or impossible. They are often conducted to evaluate the effectiveness of a treatment—perhaps a type of psychotherapy or an educational intervention. There are many different kinds of quasi-experiments, but we will discuss just a few of the most common ones in this chapter.

1. Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues in field settings*. Boston, MA: Houghton Mifflin.

# 38. One-Group Designs

*Learning Objectives*

1. Explain what quasi-experimental research is and distinguish it clearly from both experimental and correlational research.
2. Describe three different types of one-group quasi-experimental designs.
3. Identify the threats to internal validity associated with each of these designs.

## One-Group Posttest Only Design

In a **one-group posttest only design,** a treatment is implemented (or an independent variable is manipulated) and then a dependent variable is measured once after the treatment is implemented. Imagine, for example, a researcher who is interested in the effectiveness of an anti-drug education program on elementary school students' attitudes toward illegal drugs. The researcher could implement the anti-drug program, and then immediately after the program ends, the researcher could measure students' attitudes toward illegal drugs.

This is the weakest type of quasi-experimental design. A major limitation to this design is the lack of a control or comparison group. There is no way to determine what the attitudes of these students would have been if they hadn't completed the anti-drug program. Despite this major limitation, results from this design are frequently reported in the media and are often misinterpreted by the general population. For instance, advertisers might claim that 80% of women noticed their skin looked bright after using Brand X cleanser for a month. If there is no comparison group, then this statistic means little to nothing.

## One-Group Pretest-Posttest Design

In a **one-group pretest-posttest design**, the dependent variable is measured once before the treatment is implemented and once after it is implemented. Let's return to the example of a researcher who is interested in the effectiveness of an anti-drug education program on elementary school students' attitudes toward illegal drugs. The researcher could measure the attitudes of students at a particular elementary school during one week, implement the anti-drug program during the next week, and finally, measure their attitudes again the following week. The pretest-posttest design is much like a within-subjects experiment in which each participant is tested first under the control condition and then under the treatment condition. It is unlike a within-subjects experiment, however, in that the order of conditions is not counterbalanced because it typically is not possible for a participant to be tested in the treatment condition first and then in an "untreated" control condition.

If the average posttest score is better than the average pretest score (e.g., attitudes toward illegal drugs are more negative after the anti-drug educational program), then it makes sense to conclude that the treatment might be responsible for the improvement. Unfortunately, one often cannot conclude this with a high degree of certainty because there may be other explanations for why the posttest scores may have changed. These alternative explanations pose threats to internal validity.

One alternative explanation goes under the name of **history**. Other things might have happened between the pretest and the posttest that caused a change from pretest to posttest. Perhaps an anti-drug program aired on television and many of the students watched it, or perhaps a celebrity died of a drug overdose and many of the students heard about it.

Another alternative explanation goes under the name of **maturation**. Participants might have changed between the pretest and the posttest in ways that they were going to anyway because they are growing and learning. If it were a year long anti-drug program, participants might become less impulsive or better reasoners and this might be responsible for the change in their attitudes toward illegal drugs.

Another threat to the internal validity of one-group pretest-posttest designs is **testing**, which refers to when the act of measuring the dependent variable during the pretest affects participants' responses at posttest. For instance, completing the measure of attitudes towards illegal drugs may have had an effect on those attitudes. Simply completing this measure may have inspired further thinking and conversations about illegal drugs that then produced a change in posttest scores.

Similarly, **instrumentation** can be a threat to the internal validity of studies using this design. Instrumentation refers to when the basic characteristics of the measuring instrument change over time. When human observers are used to measure behavior, they may over time gain skill, become fatigued, or change the standards on which observations are based. So participants may have taken the measure of attitudes toward illegal drugs very seriously during the pretest when it was novel but then they may have become bored with the measure at posttest and been less careful in considering their responses.

Another alternative explanation for a change in the dependent variable in a pretest-posttest design is **regression to the mean**. This refers to the statistical fact that an individual who scores extremely high or extremely low on a variable on one occasion will tend to score less extremely on the next occasion. For example, a bowler with a long-term average of 150 who suddenly bowls a 220 will almost certainly score lower in the next game. Her score will "regress" toward her mean score of 150. Regression to the mean can be a problem when participants are selected for further study *because* of their extreme scores. Imagine, for example, that only students who scored especially high on the test of attitudes toward illegal drugs (those with extremely favorable attitudes toward drugs) were given the anti-drug program and then were retested. Regression to the mean all but guarantees that their scores will be lower at the posttest even if the training program has no effect.

A closely related concept—and an extremely important one in psychological research—is **spontaneous remission**. This is the tendency for many medical and psychological problems to improve over time without any form of treatment. The common cold is a good example. If one were to measure symptom severity in 100 common cold sufferers today, give them a bowl of chicken soup every day, and then measure their symptom severity again in a week, they would probably be much improved. This does not mean that the chicken soup was responsible for the improvement, however, because they would have been much improved without any treatment at all. The same is true of many psychological problems. A group of severely depressed people today is likely to be less depressed on average in 6 months. In reviewing the results of several studies of treatments for depression, researchers Michael Posternak and Ivan Miller found that participants in waitlist control conditions improved an average of 10 to 15% before they received any treatment at all (Posternak & Miller, 2001)[1]. Thus one must generally be very cautious about inferring causality from pretest-posttest designs.

A common approach to ruling out the threats to internal validity described above is by revisiting the research design to include a control group, one that does not receive the treatment effect. A control group would be subject to the same

1. Posternak, M. A., & Miller, I. (2001). Untreated short-term course of major depression: A meta-analysis of studies using outcomes from studies using wait-list control groups. *Journal of Affective Disorders*, 66, 139–146.

threats from history, maturation, testing, instrumentation, regression to the mean, and spontaneous remission and so would allow the researcher to measure the actual effect of the treatment (if any). Of course, including a control group would mean that this is no longer a one-group design.

*Does Psychotherapy Work?*

Early studies on the effectiveness of psychotherapy tended to use pretest-posttest designs. In a classic 1952 article, researcher Hans Eysenck summarized the results of 24 such studies showing that about two thirds of patients improved between the pretest and the posttest (Eysenck, 1952)[2]. But Eysenck also compared these results with archival data from state hospital and insurance company records showing that similar patients recovered at about the same rate *without* receiving psychotherapy. This parallel suggested to Eysenck that the improvement that patients showed in the pretest-posttest studies might be no more than spontaneous remission. Note that Eysenck did not conclude that psychotherapy was ineffective. He merely concluded that there was no evidence that it was, and he wrote of "the necessity of properly planned and executed experimental studies into this important field" (p. 323). You can read the entire article here:

http://psychclassics.yorku.ca/Eysenck/psychotherapy.htm

Fortunately, many other researchers took up Eysenck's challenge, and by 1980 hundreds of experiments had been conducted in which participants were randomly assigned to treatment and control conditions, and the results were summarized in a classic book by Mary Lee Smith, Gene Glass, and Thomas Miller (Smith, Glass, & Miller, 1980)[3]. They found that overall psychotherapy was quite effective, with about 80% of treatment participants improving more than the average control participant. Subsequent research has focused more on the conditions under which different types of psychotherapy are more or less effective.

## Interrupted Time Series Design

A variant of the pretest-posttest design is the **interrupted time-series design**. A time series is a set of measurements taken at intervals over a period of time. For example, a manufacturing company might measure its workers' productivity each week for a year. In an interrupted time series-design, a time series like this one is "interrupted" by a treatment. In one classic example, the treatment was the reduction of the work shifts in a factory from 10 hours to 8 hours (Cook & Campbell, 1979)[4]. Because productivity increased rather quickly after the shortening of the work shifts, and because it remained elevated for many months afterward, the researcher concluded that the shortening of the shifts caused the increase in productivity. Notice that the interrupted time-series design is like a pretest-posttest design in that it

2. Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, 16, 319–324.
3. Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore, MD: Johns Hopkins University Press.
4. Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues in field settings*. Boston, MA: Houghton Mifflin.

includes measurements of the dependent variable both before and after the treatment. It is unlike the pretest-posttest design, however, in that it includes multiple pretest and posttest measurements.

Figure 8.1 shows data from a hypothetical interrupted time-series study. The dependent variable is the number of student absences per week in a research methods course. The treatment is that the instructor begins publicly taking attendance each day so that students know that the instructor is aware of who is present and who is absent. The top panel of Figure 8.1 shows how the data might look if this treatment worked. There is a consistently high number of absences before the treatment, and there is an immediate and sustained drop in absences after the treatment. The bottom panel of Figure 8.1 shows how the data might look if this treatment did not work. On average, the number of absences after the treatment is about the same as the number before. This figure also illustrates an advantage of the interrupted time-series design over a simpler pretest-posttest design. If there had been only one measurement of absences before the treatment at Week 7 and one afterward at Week 8, then it would have looked as though the treatment were responsible for the reduction. The multiple measurements both before and after the treatment suggest that the reduction between Weeks 7 and 8 is nothing more than normal week-to-week variation.

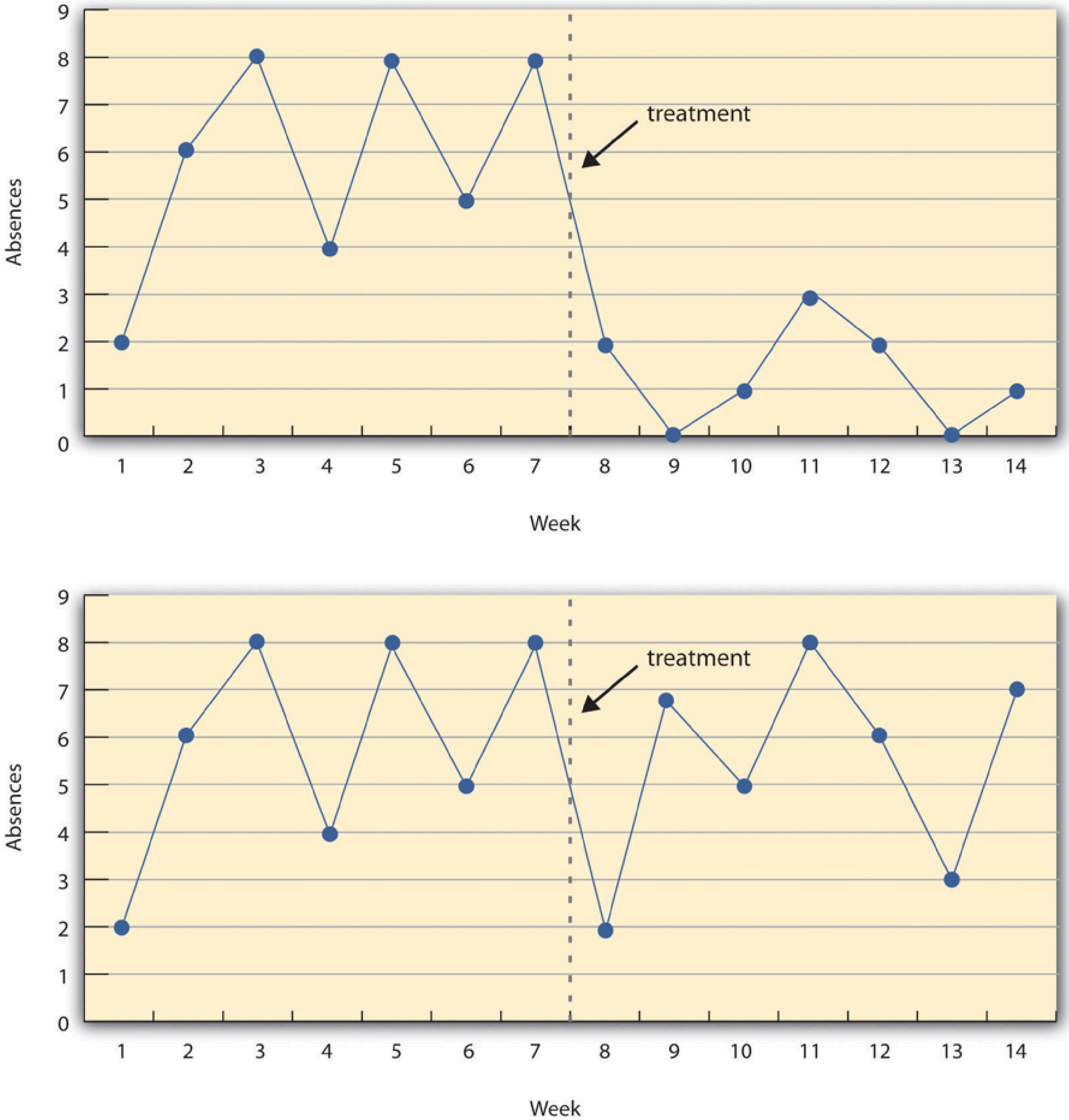


*Figure 8.1 A Hypothetical Interrupted Time-Series Design. The top panel shows data that suggest that the treatment caused a reduction in absences. The bottom panel shows data that suggest that it did not.*

# 39. Non-Equivalent Groups Designs

*Learning Objectives*

1. Describe the different types of nonequivalent groups quasi-experimental designs.
2. Identify some of the threats to internal validity associated with each of these designs.

Recall that when participants in a between-subjects experiment are randomly assigned to conditions, the resulting groups are likely to be quite similar. In fact, researchers consider them to be equivalent. When participants are not randomly assigned to conditions, however, the resulting groups are likely to be dissimilar in some ways. For this reason, researchers consider them to be nonequivalent. A **nonequivalent groups design**, then, is a between-subjects design in which participants have not been randomly assigned to conditions. There are several types of nonequivalent groups designs we will consider.

## Posttest Only Nonequivalent Groups Design

The first nonequivalent groups design we will consider is the **posttest only nonequivalent groups design.** In this design, participants in one group are exposed to a treatment, a nonequivalent group is not exposed to the treatment, and then the two groups are compared. Imagine, for example, a researcher who wants to evaluate a new method of teaching fractions to third graders. One way would be to conduct a study with a treatment group consisting of one class of third-grade students and a control group consisting of another class of third-grade students. This design would be a nonequivalent groups design because the students are not randomly assigned to classes by the researcher, which means there could be important differences between them. For example, the parents of higher achieving or more motivated students might have been more likely to request that their children be assigned to Ms. Williams's class. Or the principal might have assigned the "troublemakers" to Mr. Jones's class because he is a stronger disciplinarian. Of course, the teachers' styles, and even the classroom environments might be very different and might cause different levels of achievement or motivation among the students. If at the end of the study there was a difference in the two classes' knowledge of fractions, it might have been caused by the difference between the teaching methods—but it might have been caused by any of these confounding variables.

Of course, researchers using a posttest only nonequivalent groups design can take steps to ensure that their groups are as similar as possible. In the present example, the researcher could try to select two classes at the same school, where the students in the two classes have similar scores on a standardized math test and the teachers are the same sex, are close in age, and have similar teaching styles. Taking such steps would increase the internal validity of the study because it would eliminate some of the most important confounding variables. But without true random assignment of the students to conditions, there remains the possibility of other important confounding variables that the researcher was not able to control.

## Pretest-Posttest Nonequivalent Groups Design

Another way to improve upon the posttest only nonequivalent groups design is to add a pretest. In the **pretest-posttest nonequivalent groups design** there is a treatment group that is given a pretest, receives a treatment, and then is given a posttest. But at the same time there is a nonequivalent control group that is given a pretest, does not receive the treatment, and then is given a posttest. The question, then, is not simply whether participants who receive the treatment improve, but whether they improve *more* than participants who do not receive the treatment.

Imagine, for example, that students in one school are given a pretest on their attitudes toward drugs, then are exposed to an anti-drug program, and finally, are given a posttest. Students in a similar school are given the pretest, not exposed to an anti-drug program, and finally, are given a posttest. Again, if students in the treatment condition become more negative toward drugs, this change in attitude could be an effect of the treatment, but it could also be a matter of history or maturation. If it really is an effect of the treatment, then students in the treatment condition should become more negative than students in the control condition. But if it is a matter of history (e.g., news of a celebrity drug overdose) or maturation (e.g., improved reasoning), then students in the two conditions would be likely to show similar amounts of change. This type of design does not completely eliminate the possibility of confounding variables, however. Something could occur at one of the schools but not the other (e.g., a student drug overdose), so students at the first school would be affected by it while students at the other school would not.

Returning to the example of evaluating a new measure of teaching third graders, this study could be improved by adding a pretest of students' knowledge of fractions. The changes in scores from pretest to posttest would then be evaluated and compared across conditions to determine whether one group demonstrated a bigger improvement in knowledge of fractions than another. Of course, the teachers' styles, and even the classroom environments might still be very different and might cause different levels of achievement or motivation among the students that are independent of the teaching intervention. Once again, differential history also represents a potential threat to internal validity. If asbestos is found in one of the schools causing it to be shut down for a month then this interruption in teaching could produce a difference across groups on posttest scores.

If participants in this kind of design are randomly assigned to conditions, it becomes a true between-groups experiment rather than a quasi-experiment. In fact, it is the kind of experiment that Eysenck called for—and that has now been conducted many times—to demonstrate the effectiveness of psychotherapy.

## Interrupted Time-Series Design with Nonequivalent Groups

One way to improve upon the interrupted time-series design is to add a control group. The **interrupted time-series design with nonequivalent groups** involves taking a set of measurements at intervals over a period of time both before and after an intervention of interest in two or more nonequivalent groups. Once again consider the manufacturing company that measures its workers' productivity each week for a year before and after reducing work shifts from 10 hours to 8 hours. This design could be improved by locating another manufacturing company who does not plan to change their shift length and using them as a nonequivalent control group. If productivity increased rather quickly after the shortening of the work shifts in the treatment group but productivity remained consistent in the control group, then this provides better evidence for the effectiveness of the treatment.

Similarly, in the example of examining the effects of taking attendance on student absences in a research methods course, the design could be improved by using students in another section of the research methods course as a control group. If a consistently higher number of absences was found in the treatment group before the intervention, followed by a sustained drop in absences after the treatment, while the nonequivalent control group showed consistently high absences across the semester then this would provide superior evidence for the effectiveness of the treatment in reducing absences.

## Pretest-Posttest Design With Switching Replication

Some of these nonequivalent control group designs can be further improved by adding a switching replication. Using a **pretest-posttest design with switching replication design,** nonequivalent groups are administered a pretest of the dependent variable, then one group receives a treatment while a nonequivalent control group does not receive a treatment, the dependent variable is assessed again, and then the treatment is added to the control group, and finally the dependent variable is assessed one last time.

As a concrete example, let's say we wanted to introduce an exercise intervention for the treatment of depression. We recruit one group of patients experiencing depression and a nonequivalent control group of students experiencing depression. We first measure depression levels in both groups, and then we introduce the exercise intervention to the patients experiencing depression, but we hold off on introducing the treatment to the students. We then measure depression levels in both groups. If the treatment is effective we should see a reduction in the depression levels of the patients (who received the treatment) but not in the students (who have not yet received the treatment). Finally, while the group of patients continues to engage in the treatment, we would introduce the treatment to the students with depression. Now and only now should we see the students' levels of depression decrease.

One of the strengths of this design is that it includes a built in replication. In the example given, we would get evidence for the efficacy of the treatment in two different samples (patients and students). Another strength of this design is that it provides more control over history effects. It becomes rather unlikely that some outside event would perfectly coincide with the introduction of the treatment in the first group and with the delayed introduction of the treatment in the second group. For instance, if a change in the weather occurred when we first introduced the treatment to the patients, and this explained their reductions in depression the second time that depression was measured, then we would see depression levels decrease in both the groups. Similarly, the switching replication helps to control for maturation and instrumentation. Both groups would be expected to show the same rates of spontaneous remission of depression and if the instrument for assessing depression happened to change at some point in the study the change would be consistent across both of the groups. Of course, demand characteristics, placebo effects, and experimenter expectancy effects can still be problems. But they can be controlled for using some of the methods described in Chapter 5.

## Switching Replication with Treatment Removal Design

In a basic pretest-posttest design with switching replication, the first group receives a treatment and the second group receives the same treatment a little bit later on (while the initial group continues to receive the treatment). In contrast, in a **switching replication with treatment removal design**, the treatment is removed from the first group when it is added to the second group. Once again, let's assume we first measure the depression levels of patients with depression and students with depression. Then we introduce the exercise intervention to only the patients. After they have been exposed to the exercise intervention for a week we assess depression levels again in both groups. If the intervention is effective then we should see depression levels decrease in the patient group but not the student group (because the students haven't received the treatment yet). Next, we would remove the treatment from the group of patients with depression. So we would tell them to stop exercising. At the same time, we would tell the student group to start exercising. After a week of the students exercising and the patients not exercising, we would reassess depression levels. Now if the intervention is effective we should see that the depression levels have decreased in the student group but that they have increased in the patient group (because they are no longer exercising).

Demonstrating a treatment effect in two groups staggered over time and demonstrating the reversal of the treatment effect after the treatment has been removed can provide strong evidence for the efficacy of the treatment. In addition to providing evidence for the replicability of the findings, this design can also provide evidence for whether the treatment continues to show effects after it has been withdrawn.

# 40. Key Takeaways and Exercises

## *Key Takeaways*

- Quasi-experimental research involves the manipulation of an independent variable without the random assignment of participants to conditions or counterbalancing of orders of conditions.
- There are three types of quasi-experimental designs that are within-subjects in nature. These are the one-group posttest only design, the one-group pretest-posttest design, and the interrupted time-series design.
- There are five types of quasi-experimental designs that are between-subjects in nature. These are the posttest only design with nonequivalent groups, the pretest-posttest design with nonequivalent groups, the interrupted time-series design with nonequivalent groups, the pretest-posttest design with switching replication, and the switching replication with treatment removal design.
- Quasi-experimental research eliminates the directionality problem because it involves the manipulation of the independent variable. However, it does not eliminate the problem of confounding variables, because it does not involve random assignment to conditions or counterbalancing. For these reasons, quasi-experimental research is generally higher in internal validity than non-experimental studies but lower than true experiments.
- Of all of the quasi-experimental designs, those that include a switching replication are highest in internal validity.

## *Exercises*

- Practice: Imagine that two professors decide to test the effect of giving daily quizzes on student performance in a statistics course. They decide that Professor A will give quizzes but Professor B will not. They will then compare the performance of students in their two sections on a common final exam. List five other variables that might differ between the two sections that could affect the results.
- Discussion: Imagine that a group of obese children is recruited for a study in which their weight is measured, then they participate for 3 months in a program that encourages them to be more active, and finally their weight is measured again. Explain how each of the following might affect the results:
    - regression to the mean
    - spontaneous remission
    - history
    - maturation