His popularity ratings dropped again a short time after each event, but because of the many before and after measurements, we have considerable confidence that it was the terrorist attacks and the war that temporarily increased the president's popularity.

The Multiple Time-Series Design. As noted, both of the preceding designs suffer from the validity threat of history. The addition of a control group to that time-series design controls this last remaining threat. The Multiple Time-Series Design (see Design Q3 in Table 10.4) illustrates this. Design Q3 controls all the threats to internal validity, including history. Events other than X should affect both groups equally, so history is not a rival explanation for differences between the groups after the experimental group has been exposed to X. A design that uses a control group without random assignment might be thought of as being suspect on the threat of selection. This is less of a problem than with preexperimental designs, however, because the series of before measures affords ample opportunity to see how similar the comparison groups are.

Data from Design Q3 are plotted and analyzed the same way as for the two preceding designs. This time, however, the two lines on the graph represent the experimental group and the control group. The patterns of the two groups are compared to see what effect X produced. Figure 10.2 illustrates this by assessing the impact of "stand your ground" laws on justifiable homicide rates. "Stand your ground" laws basically relax the

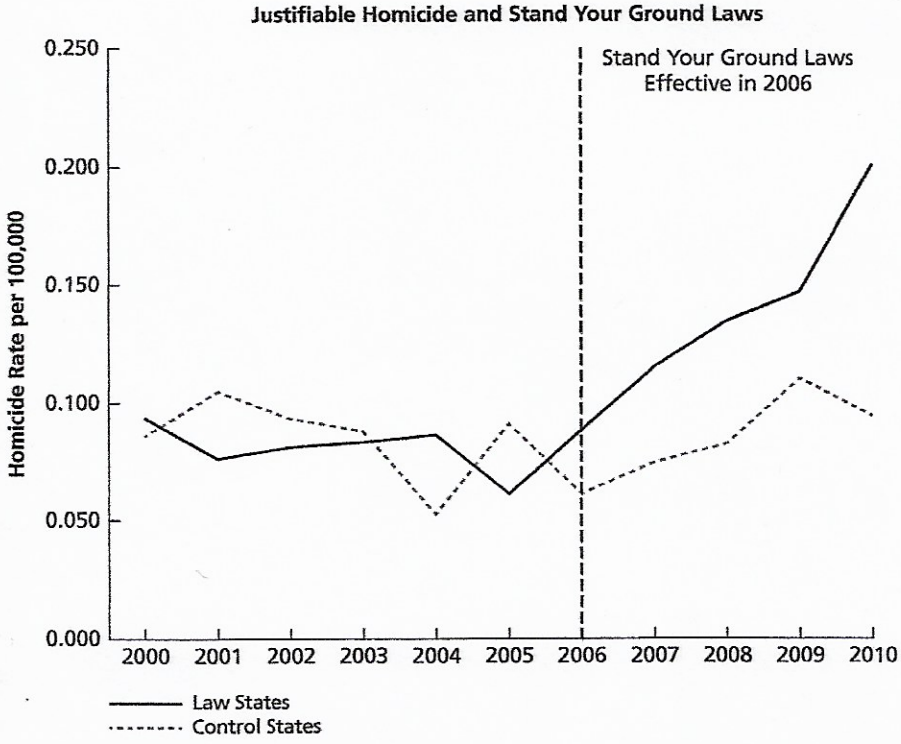

FIGURE 10.2 Example of Multiple Time-Series Data: Justifiable Homicide Rates for States With "Stand Your Ground" Laws (Law States) and States Without Such Laws (Control States), 2000–2010.

SOURCE: Federal Bureau of Investigation, *Uniform Crime Reports*, various years.

conditions under which a person can legally use lethal force when under a perceived threat. Under such laws, a person is no longer required to try to get away from the threat before using lethal force. Figure 10.2 compares states that passed such laws in 2006 with states that did not have such laws during that period. The data in the figure suggest that rates of justifiable homicide increase when jurisdictions have such laws: In states where these laws became effective in 2006, the rates began to increase right after the laws came into effect and had not increased for some years before that. The control group (states without such laws) shows steady justifiable homicide rates through the whole period under study. The control group protects against the validity threat of history, because any events (other than the "stand your ground" laws) that might have affected justifiable homicide rates should have done so in the states without such laws as well.

Quasi-experimental designs have considerable potential for human service research, much of which involves repeated contact with the same clients. Records of clients' situations and progress are routinely kept, so the repeated measures that time-series designs require may be easy to obtain. Various treatments or programs used with clients constitute the experimental condition. Plotting the data as suggested for time-series analysis would reveal whatever impacts those procedures produced. (We will discuss some of these topics and the intricacies of such designs in more detail in Chapter 11.)

Modes of Data Collection in Experiments

A good grasp of experiments in research makes one thing clear: Experiments are defined by a logical and orderly way of collecting data rather than by the form those data take. The research methods discussed in Chapters 7 to 9 are defined by the type of data collected: Surveys collect data in the form of verbal reports, available data uses archival records, and field research uses direct observation. Experiments, on the other hand, can collect data through all three of those sources—surveys, available data, and direct observation. Regardless of the kind of data collected, what makes a research design an experiment is the presence of an experimental group, a control group, random assignment to conditions, and the other factors described in this chapter.

Some experiments, then, measure variables by getting people's responses to survey questions. The study discussed earlier concerning the impact of social class and race on gender stereotypes was of this sort: People were asked to state what society's stereotypes about different types of people were (Landrine, 1985). The Minneapolis study of domestic violence and its replications described in Research in Practice 10.2 relied partially on interviews with victims to measure variables, and also used available data in the form of arrest data (Sherman, 1992).

Direct observation also is a way of collecting data in experiments, as illustrated by a quasi-experimental field study assessing child care program quality (Ridley, McWilliam, & Oates, 2000). In this case, the researchers compared the quality of care delivered by two different licensing levels of child care centers in North Carolina. A key outcome variable was "engagement," which was defined as the amount of time children spent interacting with the environment in a developmentally and contextually appropriate manner. Observers used a systematic process to code group engagement. They simply counted the number of children who were visible in one pass and then made a second pass, counting the number of children who were nonengaged (e.g., wandering aimlessly, crying, or fighting). The percentage of engaged children was then computed and used as the dependent variable.

Technological advances in recording devices, such as video recorders and computers, have enhanced the capacity of researchers to collect data by observation. The Information Technology in Research section of this chapter explores some of the tools available to aid researchers in gathering high-quality observational data for use in experiments.

In designing an experiment, then, a researcher also has to consider the issues involved in the particular kind of data collection being used. If data will be collected for an experiment through a survey, for example, all the issues of question design, nonresponse, and others discussed in Chapter 7 must be addressed. This chapter focuses on the logic behind experiments and leaves the issues related to the various forms of data collection to their respective chapters.

EXTERNAL VALIDITY

In experiments, researchers make something happen that would not have occurred naturally. In laboratory experiments, for example, they construct a social setting that they believe simulates what occurs in the everyday world. In addition to its artificiality, the setting typically is simpler than what occurs naturally, because researchers try to limit the number of social or psychological forces operating so they can observe more clearly the influence of the independent variables on the dependent variable. Even field experiments, which are considerably more natural than those in the laboratory, involve a manipulation of an independent variable by the researcher—that is, a change in the scene that would not have occurred without the researcher's intervention.

This "unnaturalness" raises a validity problem that differs from internal validity as discussed earlier in this chapter. **External validity** concerns the extent to which causal inferences made in an experiment can be generalized to other times, settings, or groups of people (Shadish, Cook, & Campbell, 2002). The basic issue is whether an experiment is so simple, so contrived, or in some other way so different from the everyday world that what we learn from it does not apply to the natural world. This problem is a part of the difficulty in generalizing findings from samples to populations (discussed in Chapter 6). With experiments, however, some special problems arise that are not found with sampling in other types of research methods. Indeed, resolving the problem of external validity can be difficult, and is less straightforward than is the case with internal validity. Campbell and Stanley (1963, p. 17) succinctly state the problem:

> Whereas the problems of internal validity are solvable within the limits of the logic of probability statistics, the problems of external validity are not logically solvable in any neat, conclusive way. Generalization always turns out to involve extrapolation into a realm not represented in one's sample.

We will review four major threats to external validity as well as ways that these threats can be reduced.

Reactive Effects of Testing

In any design using pretesting, the possibility exists that experiencing the pretest can alter the subjects' reactions to the independent variable. For example, items on a paper-and-pencil pretest measure might make subjects more or less responsive to the independent variable than they would have been without exposure to those items. The problem this raises for generalizability is not difficult to understand: The populations to which we wish to generalize findings are composed of people who have not been pretested. Therefore, if the subjects are affected by the pretest, our findings may not accurately generalize to the unpretested population.

Whether reactive effects of testing are a threat to generalizability depends on the variables that are involved in the experiment and the nature of the measurement process used. Paper-and-pencil measures may be quite reactive; unobtrusive observation is likely not to be reactive. When a researcher plans an experiment, it is important to consider whether the reactive effects of testing are likely to be a problem. If so, then it is desirable to choose a research design that does not call for pretesting, such as Design T3, or one that includes groups that are not pretested, such as the Solomon Design. With the latter, it is possible to measure the extent of any pretesting effects.

Unrepresentative Samples

As emphasized in Chapter 6, the representativeness of the people who are studied in any form of research is crucial to the issue of generalizability. Unfortunately, it often is difficult to experiment on truly representative samples of any known population. Often, experimental subjects are volunteers, people who are enticed in some way to participate, or people who happen to be available to the researcher. The implications of this for generalizing experimental findings are quite serious. For example, repeated studies over the decades have demonstrated that people who volunteer to participate in experiments often differ in systematic ways from the people in the population from which they are drawn. The *exact* nature of the differences varies from one study to another: In some cases, volunteers are better educated, from a higher social class, female (or, in other cases, male), have a greater need for social approval, or have a higher achievement motivation (Boynton, 2003; Rosnow, 1993; Zelenski, Rusting, & Larsen, 2003). Some studies show no such differences. Clearly, if these differences exist in a particular study, they could be related to any number of the variables that are likely to be used in human service research. Generalizing findings from such volunteer subjects could be quite hazardous. The best-designed experiments make an effort to measure such differences by comparing the characteristics of the volunteers with those of the population from which they were taken, if such comparative data exist.

Although the use of coerced subjects may reduce the differences between experimental subjects and the general population from which they are drawn, this has problems of its own as far as generalization is concerned. Coerced subjects are likely to have little interest in the experiment and may be resentful about the coercion that was used (Cox & Sipprelle, 1971). This effect has been found even where the nature of the coercion was quite mild, such as gaining extra credit in a college class for agreeing to participate in an experiment. One can reasonably assume that these effects are amplified where the level of coercion is greater, such as court-ordered participation in a treatment program under threat of incarceration. In fact, a study of a drug treatment program involving court-ordered clients found this anticipated pattern (Peyrot, 1985): The coerced clients were resentful, uncooperative, and unwilling to commit to the therapeutic objectives of the program.

The threat to external validity created by unrepresentative samples is of great importance to human service workers because we conduct most of our research efforts to evaluate the effectiveness of treatment approaches on volunteers or coerced subjects. Volunteers may make programs look good because they are interested in the treatment and motivated to change in the direction promoted by the treatment. Not surprisingly, however, apparently effective treatment approaches on volunteers often fail when applied to nonvolunteer groups. Alternatively, because of their lack of interest and resentment, coerced subjects tend to make treatments appear to be ineffective when these treatments might be effective on persons who are not coerced.

Reactive Settings

In addition to reactivity produced by testing, the experimental setting itself may lead people to behave in ways that are different from their behavior in the everyday world. One reason for this is that experimental settings contain **demand characteristics**: subtle, unprogrammed cues that communicate to subjects something about how they should behave. For example, people in experiments tend to be highly cooperative and responsive to the experimenter. In fact, the psychologist Martin Orne (1962) deliberately tried to create boring and repetitive tasks for subjects so that they would rebel and refuse to do them. One task was to perform a series of additions of random numbers. Each page required 200 additions, and each person was given 2,000 pages. After giving instructions, the experimenter told them to continue working until he returned;

hours later, the subjects were still working—it was Orne who gave up and ended the experiment! On the basis of this and other research, it has become evident that experimental settings can exercise an influence on subjects powerful enough to damage the generalizability of experimental findings. This has been labeled as the problem of the *good subject*—that is, the subject in an experiment who will do whatever the investigator asks, even to the point of confirming the experimenter's hypotheses if they are communicated to the subject (Aronson, Brewer, & Carlsmith, 1985).

Subjects' reactions are not the only way that experimental settings can be reactive. Experimenters themselves can introduce distortion into the results that reduce generalizability. Experimenters, of course, usually have expectations concerning the results of the experiment, wanting it to come out one way or another. These **experimenter's expectations** can be communicated to subjects in such a subtle fashion that neither the experimenter nor the subjects are aware that the communication has taken place (Rosenthal, 1967). A classic illustration of how subtle this can be is recounted by Graham (1977) and involves a horse, not humans. The horse was called Clever Hans because of his seeming ability to solve fairly complex arithmetic problems. Hans and his trainer toured Europe during the early 1900s, amazing audiences and becoming quite famous. Hans would stand on stage pounding out the answers to problems with his hoofs. Amazingly, Hans was hardly ever wrong. Hans would perform his feats even when his trainer was not present. Could the horse really do arithmetic? After very careful observation, it was discovered that Hans was picking up the subtle cues given off by those who asked him questions. As Hans approached the correct number of hoof beats, questioners would move or shift just enough to cue the horse that it was time to stop.

The same phenomenon has been thoroughly investigated in the physician–patient relationship (Kline, 1988). When giving a patient a treatment that is known to be effective, a physician exudes confidence, and the patient picks up on this and expects to get better. The patient's hopeful mood then increases the likelihood that he or she will actually improve. When a physician, however, is giving a patient a placebo—that is, an inert substance the physician doesn't expect to work—his or her manner communicates doubt: The patient senses, without being aware of it, what the physician's expectations are and then helps create that reality. People in experiments often do the same thing. Subjects react to subtle cues given off unconsciously by an experimenter that tell them to behave in a way that the researcher wants. (The Eye on Ethics section discusses some ethical issues relating to reactivity in experiments.)

Reactive experimental settings can threaten external validity because changes in the dependent variable might be caused by demand characteristics or experimenter expectancies rather than by the independent variable. One procedure for reducing reactivity in experimental settings is to conduct the experiment in such a way that subjects are *blind*—that is, unaware of the experimental hypotheses—so this knowledge will not influence their behavior. In fact, subjects commonly are given a false rationale for what they are to do in an attempt to reduce their reactions that might interfere with the generalizability of the findings. As an ethical matter, however, subjects should be informed of the experiment's true purpose during the postexperimental debriefing session.

The surest way of controlling reactivity because of the experimenter is to run what is called a *double-blind experiment*. In a **double-blind experiment**, neither the subjects nor the experimenter knows which people are in the experimental condition and which are in the control condition. This makes it impossible for the experimenter to communicate to subjects how they ought to behave, because, for any given subject, the experimenter does not know which responses confirm the experimental hypotheses. Though simple in theory, maintaining a double-blind procedure in practice can be difficult. Another layer of personnel must be added to accomplish the assignment of subjects, sealed instructions issued to the experimenter, and the results tracked. Furthermore, all information

EYE ON ETHICS **Deception and Informed Consent in Experiments**

Two of the most common ethical issues that arise in experimental research are deception and informed consent. Deception can occur when experimental subjects are not told about some key aspects of the experiment or are actually misled about them. Most commonly, subjects might not be told about the true purposes of the research, or they might actually be told something that is not true. For example, subjects might be misled by being given some tasks or a questionnaire that are really unrelated to the experiment.

The reason this deception is necessary is that if subjects know the true purposes or the actual research hypotheses, they might behave differently. For example, they might try to help the researcher by behaving in a way that helps confirm the hypotheses. The subjects' behavior would then be a form of reactivity that throws into question the validity of the findings.

Researchers justify these deceptive practices in a number of ways. They argue that the research findings are important and the research goals could not be achieved without the deception. They also argue that the deception in most cases is rather minor in nature.

To alleviate some of the ethical concerns, researchers often have the subjects read and sign a consent form that informs them, among other things, that they may be deceived as to the true purpose of the experiment and may be asked to do some things to further that deception. Yet another way to deal with the problem of deception is to conduct a debriefing, a session following the experiment when subjects are told the complete nature of the study, including any deceptions used, and their reactions are assessed. Especially when deception is used, a debriefing is an ethical obligation of the researcher to ensure that people are in no way harmed by their participation. During debriefing, the researcher assesses whether the subject has had any kind of negative reaction to the deception or any other element of the experiment. It sometimes happens that things people read or do during an experiment produce feelings of anger, depression, or despair. The researcher's obligation is to assess these reactions and alleviate them, with professional assistance if necessary.

Ethical absolutists would argue that no deception can be justified. They would question whether subjects can truly give informed consent if they are misled about some aspects of the research. Most social scientists, however, would argue that mild levels of deception, if done carefully, are necessary and ethically acceptable.

relating to these activities must be kept from those who are actually running the experimental groups. It is easy for this structure to break down so the double-blind feature is lost; however, for true protection against such reactivity, double-blind experiments need to be used. Unfortunately, the nature of the treatment sometimes interferes with use of the double-blind approach. Because of the different activities involved, it may be obvious which condition is experimental and which is control, thus clearly preventing the use of a double-blind experiment.

Multiple-Treatment Interference

In an experiment with more than one independent variable, the particular combination and ordering of experimental treatments may produce change in the dependent variable. If this same combination and this same ordering do not occur outside the experimental setting, however, the findings from the experiment cannot be generalized. Suppose, for example, that an experiment calls for subjects to experience four independent variables in succession. Furthermore, suppose that the last variable in the series appears to produce an interesting effect. Could the effect of that fourth variable be safely generalized on the basis of the experimental findings? The answer is no. The subjects in the experiment have experienced the three other independent variables first, which might in itself have affected the way that they reacted to the last variable. If people outside the experimental setting do not experience all the variables in sequence, then generalization concerning the fourth variable is risky. The problem of multiple-treatment interference is similar to reactive effects of testing in that the subjects in the experiment experience something that the people in the population at large do not.

The threat of multiple-treatment interference can be effectively eliminated through the use of complex designs in which the different experimental groups experience the independent variables in every possible sequence. If a given variable produces a consistent effect regardless of the ordering of the variables, then multiple-treatment interference is not a threat to the generalizability of the findings. An alternative is to isolate the variable of interest in a multiple-treatment experiment and then conduct a follow-up experiment with that variable as the only treatment. If the follow-up experiment produces an effect similar to that found when it was part of a series of treatments, multiple-treatment interference can be ruled out.

In designing experiments, especially those to be conducted in laboratory settings, problems of external validity need to be considered. Properly designed experiments can offer researchers substantial confidence in generalizing from an experimental sample to other settings or groups. Efforts to enhance external validity may, however, affect our ability to achieve internal validity, which brings us to the importance of replication for external validity.

Enhancing the External Validity of Experiments

Although both internal and external validity are important to experimental research, a tension exists between the two. On the one hand, internal validity is enhanced through greater control. Consequently, the researcher seeking internal validity is attracted to the laboratory experiment and precise testing. Yet, we have just seen that the reactive effect of testing is a threat to external validity. Seeking control, the researcher may use a homogeneous sample to avoid the confounding effects of other variables. In the effort to achieve internal validity, the researcher might, for example, use a sample of 18-year-old white males. Although the effects of race, sex, and age are now controlled, the threat to external validity from unrepresentative samples increases. We have suggested the use of complex designs to reduce the threat to external validity of multiple-treatment interference. Complex designs, however, are more difficult to implement in a way that retains the integrity of the research design. Thus there appears to be a dilemma between seeking internal validity on the one hand and external validity on the other.

Another way to approach the problem of external validity is, when possible, to conduct field experiments rather than laboratory experiments. Although issues of external validity do arise in field experiments, field experiments on the whole have greater external validity because they are conducted in the real world. This means that the results are more clearly generalizable to some settings in the real world, at least to settings that are similar to the one in which the field study was conducted.

The basic solution to this dilemma, however, is not to seek an answer to both internal validity and external validity in a single study. Ultimately, external validity is established through replication (Shadish, Cook, & Campbell, 2002). Confidence in generalizing from experimental findings increases as other researchers test and support the same hypotheses in a variety of settings with a variety of designs. Often, this takes the form of initial testing in the laboratory under ideal conditions to see if the hypothesis is supported at all. Then the same hypothesis can be tested under the less-than-ideal conditions of the field. Through replication, the dual objectives of internal and external validity can be achieved.

Lack of Minority Participation and Analysis

Depending on the setting in which an experiment is conducted, women and minorities may find themselves underrepresented in laboratory or field experiments—a problem that can affect both external and internal validity. Researchers often choose people who are easily accessible to them as subjects for experiments. For basic research in psychology and sociology, for example,

students in introductory psychology or sociology classes often are selected as research subjects. The sex ratio of college students is fairly even: 46 percent male, and 54 percent female. African Americans constitute about 11 percent of all college students but account for 13 percent of the total U.S. population (U.S. Census Bureau, 2011). Thus experiments using a representative sample of college students will tend to underrepresent African Americans, Hispanics, and some other minorities. In addition, this underrepresentation is more severe at some colleges than at others: Some major research universities, even today, have only 1 percent or 2 percent black enrollment. So whenever experimental subjects are selected from a setting, care must be taken to ensure representation of minorities in a sample if the minorities are to be included in the generalizations made from the data. Otherwise, it should be mentioned that no special procedures were used in the sampling to enhance minority involvement.

A related problem is what one sociologist calls *gender insensitivity*: "ignoring gender as an important social variable" (Eichler, 1988, p. 66). In experiments, this can happen when no mention is made of the sex ratio of the subjects in the experiment, or when data are not analyzed separately for each gender. Eichler reports one issue of a psychology journal in which the only article to mention the sex of the subjects was one that used rhesus monkeys; none of those using human subjects did so, despite the fact that practically all those articles focused on variables, such as perception, verbal ability, and learning, on which people's gender might well have an influence.

A final problem relating to minority participation in experiments is failure to consider the gender or minority status of all participants in the experiment. In addition to the researcher and the subject, this might include interviewers, confederates of the experimenter, and any others who interact with the subjects during data collection. We know, for example, that people of the same gender interact differently with people of the opposite gender. Thus, a male interviewer responds differently when asking questions of a female subject than when interviewing a male subject. In the former case, he might be friendlier, more attentive, or more engaging without even realizing it, and this can influence the subject's response to questions.

These threats to external validity can be reduced by reporting the sex or minority status of the various people who are involved in experiments and analyzing the data with sex or minority status as an experimental or control variable. There are, of course, reasons for a homogeneous sample in an experiment—namely, to reduce extraneous variation when trying to establish a causal relationship between independent and dependent variables. If the results are to be generalized to all racial and ethnic groups and both sexes, however, then experiments need to be replicated using subjects from these groups.

ASSESSMENT OF EXPERIMENTS

Advantages

Inference of Causality. The major advantage of experimental research is that it places us in the most advantageous position from which to infer causal relationships between variables. (Recall from Chapter 2 that causality can never be directly observed. Rather, we infer that one thing caused another by observing changes in the two things under the appropriate conditions.) A well-designed and well-controlled experiment puts us in the strongest position to make that causal inference, because it enables us to control the effect of other variables, which raises our confidence that the independent variable is, indeed, bringing about changes in the dependent variable. Experiments also permit us to establish the time sequence necessary for inferring causality. We can measure the dependent variable to see if it has changed *since* the experimental manipulation. With other research methods, such as the survey, it often is not possible to directly observe whether the

changes in one variable preceded or followed changes in another variable.

Control. Experimenters are not limited by the variables and events that happen to occur naturally in a particular situation. Rather they can decide what variables to study, what values those variables will take, and what combination of variables to include. In other words, researchers can create precisely what their hypotheses suggest are important. This is in sharp contrast to other research methods, such as the use of available data (see Chapter 8) or observational techniques (see Chapter 9), in which hypotheses need to be reshaped to fit the existing data or observations.

The Study of Change. Many experimental designs are longitudinal, which means they are conducted over a period of time, with measurements taken at more than one time. This makes it possible to study changes over time. Many other research methods, such as surveys, tend to be cross-sectional—they are like snapshots taken at a given point in time. In cross-sectional studies, we can ask people how things have changed over time, but we cannot *directly observe* that change.

Costs. In some cases, experiments—especially those conducted in laboratory settings—can be considerably cheaper than other research methods. Because of the element of control, the sample size can be smaller, and this saves money. Costly travel expenses and interviewer salaries found in some surveys also are eliminated. Some field experiments, however, can be expensive, because they may require interviewing—possibly more than once—to assess changes in dependent variables.

Disadvantages

Inflexibility. Experimental designs generally require that the treatment (or the independent variable) be well developed, and that the same treatment be applied to all cases in the experimental group. If the nature of the independent variable changes as the experiment progresses, this is extraneous variation, which makes it difficult to say what changed the dependent variable. Because of this inflexibility, experimental designs are not suitable during the early stages of research when treatments may not be well developed. This requirement for consistently measuring the independent variable is especially true of large-scale, longitudinal experiments, which are best suited for clearly designed treatments (Rossi, Lipsey, & Freeman, 2004).

Randomization Requirement. Human service organizations may be unwilling to accept random assignment to treatment and control conditions, despite the value of this technique for controlling extraneous effects. This is especially so when the design calls for a control group to receive no treatment at all. Organizations may have ethical concerns about randomization, or they may want to be sure that they "get something" for participating in a study, and thus are reluctant to serve as the control condition. In addition, with field experiments it sometimes is impossible to find or create a control group that is truly comparable to the experimental groups; thus the true experimental designs are ruled out. Despite the fact that randomization was achieved, both Research in Practice illustrations in this chapter emphasize the considerable difficulty in carrying out randomization.

Artificiality. The laboratory setting in which some experiments are conducted is an artificial environment created by the investigator, which raises questions of external validity. We often do not know what the relationship is between this artificial setting and the real world in which people live. We cannot be sure that people behave in the same fashion on the street or in the classroom as they do in the laboratory. Even field experiments of human service interventions are, to an extent, vulnerable to this criticism, especially when they create conditions that never exist naturally. For

INFORMATION TECHNOLOGY IN RESEARCH **Computer Applications in Experimental Research**

In defining "experiments," we stated that the essence of experimentation is a *controlled method of observation* in which the value of one or more independent variables is changed to assess its causal effect on one or more dependent variables. Researchers endeavor to eliminate variation resulting from any variables other than the independent variable. In this effort to reduce extraneous variation, computer technology can be helpful.

One strategy to reduce extraneous variation and produce higher-quality data is to have research participants themselves record data on some type of electronic device. Such research results in more accurate and timely data and is preferred by research participants over paper and pencil data collection methods (Haller et al., 2009; Lane et al., 2006). The computer also may incorporate a signaling function that alerts participants when it is time to record data (Tejani, Dresselhaus, & Weinger, 2010; Fahrenberg, 1996). For example, one research team employed handheld computer devices to study whether smokers found it more rewarding to smoke when they chose to rather than when told to by researchers. Participants completed short, 11-question surveys using Palm m505 handheld computer devices to evaluate their moods after smoking cigarettes at times of their choosing on three separate days. Participants later took the same survey after smoking cigarettes at times chosen by the researchers. The handheld computers signaled when participants should smoke. Results showed that participants reported it less rewarding (as measured by poorer moods, less reductions in cravings, and less improvement in overall feelings) when they smoked at prescribed times than when they smoked when they chose to (Catley & Grobe, 2008).

Gravlee et al. (2006) argue that handheld computers are also useful in field research for recording systematic observations of social and physical environments. They describe their use of handheld computers in completing Healthy Environments Partnership's Neighborhood Observational Checklists, a 140-item survey assessing community resources. Handheld computer devices, they argue, are preferable to paper and pencil formats because they are easy to use, allow for multiple users, create customized surveys based on variable sizes (e.g., the software cycles through series of street-level questions based on the number of streets in an area), improve data quality, and make exporting data much easier.

The number of available tools for data collection is expanding. One example is Palm-tungsten E2, which is popular because data can be easily downloaded from the handheld device to Access-Microsoft or Stata tables on a laptop or desktop computer. Another device, HanDBase professional is popular because of the ease with which filter and contingency questions (see Chapter 7) can be created as well as items that require pull-down menus. Surveys with response buttons, checkboxes, pop-up lists, and automatic date and number entries can also be created relatively easily with HanDBase (Haller et al., 2009).

example, the recipients of social services who are the subjects of research may be treated better by the research team than they would be during their everyday contacts with human service agencies. Under the watchful eye of a research team, much greater care may be taken in the form of monitoring and supervision to ensure that the treatment being delivered is true to the theoretical model than might be taken when the same intervention is delivered as a routine part of an agency's service.

Experimenter Effects. As discussed, artificiality of experimental settings in comparison to the real world exists, and there is another side to this coin: The experimental setting itself is a real world—that is, a social occasion in which social norms, roles, and values exist and shape people's behavior. Social processes arising within the experimental setting—and not a part of the experimental variables—may shape the research outcome. Thus, changes in the dependent variable may result from the demand characteristics of experiments, such as repeated

measurement, or from the impact of experimenter expectations, which are factors that obviously do not influence people in the everyday world.

Generalizability. The logistics involved in managing an experiment often necessitate that experimenters use small samples. Furthermore, the goal of holding measurement error to a minimum often compels researchers to make these samples as homogeneous as possible. Despite the advantages of random sampling, experimenters frequently must settle for availability samples to obtain subjects who will participate in the study. The net result is that experiments often are conducted on small, homogeneous, availability samples, and it may be difficult to know to whom such results can be generalized. It often is assumed that representative sampling is not as essential in experiments as in other types of research because of the randomization and control procedures that are used. Although these help, we must still be cautious regarding the population to which we make such generalizations.

Timeliness of Results. Although experiments may produce the strongest evidence of causation, they often are time-consuming to conduct. The urgency of policy decisions may preclude using experimentation simply because it takes too long to generate funding for the project, design the research, and carry it out. If the results do not come in quickly enough, they may not be considered in the debate over a rapidly developing policy (Rossi, Lipsey, & Freeman, 2004).

REVIEW AND CRITICAL THINKING

Main Points

- Experiments are a controlled method of observation in which the value of one or more independent variables is changed to assess the causal effect on one or more dependent variables.
- Because of the control that experiments afford, they are the surest method of discovering causal relationships among variables.
- Major elements of control in experiments include control variables, control groups, matching, and randomization.
- Internal validity refers to whether the independent variable does, in fact, produce the effect that it appears to produce on the dependent variable.
- Numerous conditions may threaten the internal validity of experiments, but using true experimental designs can control all of them.
- Quasi-experimental designs are useful for bringing much of the control of an experiment to nonexperimental situations.
- External validity is the degree to which experimental results can be generalized beyond the experimental setting.
- Threats to external validity generally are not controllable through design, but are more difficult and less straightforward to control than are threats to internal validity.
- Double-blind experiments, in which neither the subjects nor the experimenters know which groups are in the experimental or control condition, are used to control both experimental demand characteristics and experimenter bias.
- To avoid race, ethnicity, or gender insensitivity, the minority composition of experimental groups should be reported, consideration should be given to analyzing the data by controlling for minority impact, and the minority status of all experimental participants should be considered.
- An understanding of experimental procedures can contribute to practice effectiveness as well as research endeavors.