Chapter 11

The Research Design

Outline

11-1 Introduction
11-2 The Research Design: An Overview
   11-2a Purpose
   11-2b Factors Affecting the Choice of Design
11-3 Preexperimental Designs
11-4 Experimental Designs
   11-4a The Classic Experimental Design
   11-4b The Basics of an Experimental Research Design
   11-4c Other Types of Experimental Designs
11-5 Nonexperimental Designs
   11-5a Nonrandomized Control Group Design
   11-5b Time Series without Control Group Design
   11-5c Time Series with Control Group Design
11-6 Other Types of Research Designs
   11-6a The Case Study
   11-6b Field Research
11-7 Obstacles to a Successful Research Effort
   11-7a Internal Validity and Extrinsic Obstacles
   11-7b Internal Validity and Intrinsic Threats
   11-7c External Validity
11-8 A Comparison of Research Designs
Chapter Summary
Chapter Quiz
Suggested Readings

Key Terms

- classic experimental design
- comparison
- control
- control group
- experimental group
- experimental mortality
- external validity
- history
- instrumentation
- internal validity
- manipulation
- matching
- maturation
- nonexperimental design
- posttest
- pretest
- regression artifact
- research design
- test group
- testing
- time series design
11-1 Introduction

So far, we have discussed the first three stages of the systematic research process. We also presented examples and exercises to enhance your comprehension. You should realize why it is important to clearly define your research problem, conduct a thorough and concise literature review, and understand concepts, variables, and hypotheses. We also hope that you know why measurement and the different levels of measurement are important to your research effort.

Now it is time to learn about the research design and the different types of research designs you have at your disposal. Research designs are important. They provide you with a plan to use when collecting, analyzing, and interpreting data you believe provides answers to your research problem. In this chapter we discuss the various research designs, the issues of internal and external validity, and the impact of extrinsic and intrinsic factors on the validity of research designs.

An understanding of this chapter will enable you to

1. Understand the purpose and importance of research designs.
2. Understand the basics of research designs.
3. Distinguish the difference between internal and external validity as they pertain to research designs.
4. Understand the components and process associated with experimental designs.
5. See the importance of nonexperimental designs.

11-2 The Research Design: An Overview

Before you test hypotheses you must develop a research design to use when collecting, analyzing, and interpreting observations and data. A research design is a plan specifying how you intend to fulfill the goals of your study. It is a logical plan for testing hypotheses.

research design: A plan that specifies how the researcher intends to fulfill the goals of the study. It is a logical plan for testing hypotheses.

11-2a Purpose

The research design serves several purposes. First, it suggests the necessary observations you need to make to provide answers to the research question. It outlines the ways you should make your observations. Second, the research design identifies the analytical and statistical procedures you will need to use when analyzing the data. A major purpose of research is to establish that the independent and dependent variables are causally related. The research design consists of four components necessary to establish this purpose: comparison, manipulation, control, and the ability to generalize findings (Frankfort-Nachmias and Nachmias 2000, 88). As we discuss in later chapters, the research design also specifies a model you can use to test the validity and significance of the statistical relationships.

11-2b Factors Affecting the Choice of Design

There are various types of research designs that you can use. The one you select depends on the following considerations.

• First, what is the purpose of your investigation? Is it to explore and describe some political phenomenon? Do you also want to explain what you have discovered? Or do you want to pursue both purposes?
• Second, what resources do you have for your research? Are the necessary data readily available? Are you impeded by time limitations such as the constraints you face when trying to complete a term paper within a school semester? Do you have financial limitations? Is a control group available? Frequently it is expensive to produce hundreds of questionnaires and travel to data sources. Often you will need to find ways to compensate for resource limitations without unduly detracting from your research endeavor.

• Third, there are threats to the internal and external validity of your research findings. Some of the designs we cover are stronger in internal validity, while others enhance the external validity of your study.

• Last, you need to remember our discussion about ethics and research. The privacy of your subjects is foremost. They place their faith in you when they agree to be a part of your research. Do not violate their privacy or their trust.

As you can see, there are several factors that affect the type of research design you use. Often they limit your choices to less than the ideal. As a result, your conclusions may be limited and somewhat imperfect.

11-3 Preexperimental Designs

Political scientists have tried to analyze individual behavior for a long time. In addition, they have used experimental designs, or some semblance thereof, to examine behavior. Prior to discussing experimental designs, however, we want to spend some time discussing preexperimental designs.

Political scientists have tried to analyze individual behavior for a long time. Many have used experimental designs, or some semblance thereof (nonexperimental designs), to achieve this purpose. Political scientists have also used preexperimental designs. These designs, however, lacked one or more of the components necessary to establish causality. **Comparison**, one of the necessary components to show causation for example, is lacking with these designs. Therefore, the designs are neither reliable nor valid. Consequently, political scientists turned to the classic experimental design as a way to enhance comparison, manipulation, and control.

11-4 Experimental Designs

Campbell and Stanley write that experimentation is “...that body of research in which variables are manipulated and their effect upon other variables observed” (Campbell and Stanley 1966). This definition implies several things. First, a behavior is studied (the dependent variable). Second, factors impacting that behavior are identified (independent variables). Third, the researcher, through manipulation, determines the effect of the independent variables on the dependent variable. A study about voting behavior, for example, might examine the impact of age, income, and education on voting. In addition, control is essential. We must know whether age, income, and education did impact voting, or whether some other factor also impacted voting. One way to control for the impact of extraneous factors is to use the classic experimental design. A classical experimental design model is illustrated in Figure 11-1.

11-4a The Classic Experimental Design

The development of the **classic experimental design** involves several important steps. The first step, **selection and assignment**, involves two comparable groups: the
test group and the control group. These groups are equivalent except that you expose the test group to the independent variable or treatment. The control group does not receive the stimulus but instead serves as a baseline for evaluating the behavior (dependent variable) of the experimental group or test group. You randomly assign units of analysis to each group. Random assignment means that you base the assignment of units on chance. Each possible unit of analysis has the same probability of assignment to either group. As such, this method posits the notion that incidental factors are the same for each group. It also assumes that each group is equivalent with respect to the dependent variable.

The second step is to pretest each group. The pretest enhances the validity of the procedure because it helps you determine the equivalency of the groups. The pretest results for each group should be similar. You can also consider pretest differences when determining the effect of the independent variable.

The third step is to treat the test group. In other words, you subject the group to the independent variable. You do not treat the control group.

The next step is to posttest each group in order to measure the dependent variable. The purpose of this step is to determine the impact of the independent variable on the behavior of the test group.

Last, you compare the measurement results to see if there are differences between the two groups. You can also compare measurement differences between pretest and posttest behaviors.

Let's summarize what we have discussed so far about the classic experimental design. The design requires you to do each of the following:

- Randomly select and assign cases to test and control groups.
- Pretest each group.
- Administer the stimulus.
- Take posttest measurements so that you can compare and make inferences from the results.

Let’s try to simplify our discussion by applying the classic design to the classroom setting. Say you enrolled in a special summer school seminar about social policy. During the introductory session, you discover that you and the other forty-nine members of your seminar have some common characteristics such as age, grade point average, and class standing. Then your professor does something strange. She randomly divides the seminar into two comparable groups and administers a test to each group. She then gives one group a selection of readings to complete during the semester. She also tells the group not to report to class until the last day of the semester. You, of course, are in the group that must come to class. In addition, your group must complete the same list of readings as the group that does not have to attend class.
The semester is intense but interesting. Your instructor has given you much to think about and, in your opinion, broadened your knowledge about social problems and government attempts to address them. The semester quickly draws to a close, and it is soon time to take the final examination. When you come to class to take the exam, you notice that the group that was not required to attend the sessions is also sitting for the exam. You also learn that they will take the same exam that you will take. It is obvious to you, however, that your score and the scores of those who attended class will be higher. After all, you had the benefit of your professor’s lectures and discussions among members of your group.

Look at this example closely. What is going on? What is the professor trying to do? Which group, if any, will score higher on the exam? While simplistic, this is an example of a classic experimental research design. All of the components are present. There is a pretest. There is an experimental group (those who continued to attend class) and a control group (those excused from regular attendance). The professor’s lectures represent the independent variable, or treatment. The final exam, of course, is the posttest.

Do the lectures enhance learning? To find out, you need to compare the test results of the two groups. If your group averaged 85 percent on the test compared to 70 percent for the other group, your professor will infer a causal relationship between class attendance and the higher scores.

How safe is the professor when she infers causation between test scores and attendance? Prior to the experiment, were the groups similar in knowledge about social problems? The classic design allows your professor to answer these questions. For example, suppose a comparison of pretest scores showed that both groups averaged approximately 55 percent. As a result your professor asserted that the two groups were comparable and equivalent. If the results were different, say 50 percent for the control group and 60 percent for the test group, she would have to admit that the groups, although comparable, were not equivalent.

11-4b The Basics of an Experimental Research Design

Before we move on to other types of research designs, we want to spend a few moments discussing the basics of the experimental research design. As implied by Campbell and Stanley’s definition of experimentation, an experimental research design consists of three components: comparison, manipulation, and control. Frankfort-Nachmias and Nachmias (2000) add a fourth component: generalization. Each component is essential in establishing a causal relationship between the independent and dependent variables. In Chapter 10, we said there are four major factors that help determine causation: covariance, time order, nonspuriousness, and theoretical basis. With an effective research design, comparison allows you to show covariation; manipulation helps establish the time order of events; and control enables you to show that the relationship is nonspurious. Generalization is our ability to extend the results of our research findings to larger populations and to different social or political settings. In other words, we are extending the boundaries of the theory we use as the basis for our research. We discuss this component when we cover the concept of external validity in Section 11-7c.

Comparison

The classic research design, through comparison, enables you to show the first requirement of causality: covariation, or correlation. In our example, the professor wanted to demonstrate a correlation between final exam scores in her social policy seminar and class attendance. In other words, class attendance is associated
with better exam scores. To test this theory, she compared the final exam scores of those who attended class with those who did not. If her hypothesis was correct, her comparison would reveal higher scores for those who attended class. She could also infer from the results that her lectures enhanced learning and test scores. She could verify this hypothesis by comparing the pretest and posttest scores of those who attended class to see if the lectures made a difference. In sum, to assess covariation in our example, you can evaluate students’ knowledge about social programs before and after the seminar lectures. Or you can compare the scores of the group that attended the lectures with the scores of the group that did not benefit from the lectures. In one instance, you compare a group with itself; in the other example, you compare an experimental group with a control group.

**Manipulation**

To demonstrate causation, we must also show that a change in one variable causes a change in another variable. This can be demonstrated through the process of manipulation. It is important to remember that with causal relationships one variable is the force that determines the response of another variable. The determining, or independent, variable must occur before the response, or dependent, variable. Thus, the time order of events is another important factor when demonstrating a causal relationship.

Let’s return to our seminar example. Causality implies that if grade scores (knowledge) are enhanced by seminar attendance, then an induced change in attendance will result in a change in knowledge. You can show this by manipulating the attendance of the group members. You allow the experimental group to come to class and exclude the control group members from the lectures. What you have really done is control the time order of events. The independent variable, lecture attendance, precedes the dependent variable, knowledge about social programs. Based on a comparison of test scores, you can determine whether knowledge increased after application of the independent variable.

**Control**

A third criterion of causality, control, requires you to prove that other factors do not explain the observed association between the variables you want to study. In other words, you must prove that the relationship you find is not spurious. A basic problem in causal analysis is the elimination of alternative causal interpretations. You might satisfy the criteria of time order, however; there are several factors that might account for the fact that two variables demonstrate covariance. X might cause Y, Y might cause X, both might be caused by Z, or causation might not exist. Your task is to show that X caused Y, thereby excluding the other possibilities. You want to provide a single causal interpretation of your observed relationship by controlling for the effects of other variables. The classic and other experimental designs will help you eliminate most of these possibilities.

**11-4c Other Types of Experimental Designs**

In this section we discuss two other types of experimental designs that are offshoots of the classic experimental design: the Solomon Four-Group Design and the Posttest Only Control Group Design.

**Solomon Four-Group Design**

The Solomon Four-Group Design is an extension of the classic research design. It is a very powerful experimental approach that some methodologists believe can be
The Research Design

used to counteract short-term changes that may occur whether the treatment is effective or not. This phenomenon is called the Hawthorne effect (Welch and Comer 2001, 20). When you pretest groups, you need to determine the effect the pretest might have on the groups. A pretest may motivate the groups to perform well under experimental conditions. The pretest may amplify the effect of the experimental variable. In sum, the pretest and stimulus could interact to further change the dependent variable beyond the effects attributable to the pretest and stimulus separately. Thus, while a pretest has advantages such as providing an assessment of the time sequence as well as a basis of comparison, it can have severe reactive effects (Frankfort-Nachmias and Nachmias 2000, 104).

The inclusion of an additional test and control group in the research design addresses this possibility. Figure 11-2 is a model of the Solomon Four-Group Design. The model reveals that the additional groups do not take a pretest. This procedure enhances your analysis of the impact of the independent variable on the dependent variable. By examining the posttest scores, you can generalize the results of the experiment on the original experiment and control groups to the results experienced by the additional groups. You can also isolate the interaction impact of the stimulus and pretest.

Consider this example. You design an experiment to determine the impact that a lecture about welfare might have on student attitudes toward welfare recipients. You randomly assign the participating students to two groups. You also pretest each group so that you can determine their attitudes. You then require the test group to attend the lecture. Later, you retest each group on their attitudes toward welfare recipients. Table 11-1 depicts the results you might observe if the lecture was the only reason for a change in attitudes. Each group scored the same on the pretest, indicating equivalency between the groups. The posttest scores, however, are different. The test group scored higher than the control group. In addition, the posttest score for the control group is the same as their pretest score. This suggests the group members were not influenced by the pretest. Consequently, you can assert that the lecture was the sole reason for change in the experiment group’s scores.
Table 11-1 Lecture and Pretest Impact on Attitudes toward Welfare Recipients (No Interaction)

<table>
<thead>
<tr>
<th>Random Assignment</th>
<th>Pretest</th>
<th>Lecture</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>70</td>
<td>Yes</td>
<td>85</td>
</tr>
<tr>
<td>Control group</td>
<td>70</td>
<td>No</td>
<td>70</td>
</tr>
</tbody>
</table>

Table 11-2 Lecture and Pretest Impact on Attitudes toward Welfare Recipients (Interaction)

<table>
<thead>
<tr>
<th>Random Assignment</th>
<th>Pretest</th>
<th>Lecture</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>70</td>
<td>Yes</td>
<td>85</td>
</tr>
<tr>
<td>Control group</td>
<td>70</td>
<td>No</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 11-3 Lecture and Pretest Impact on Attitudes toward Welfare Recipients (No Interaction) Solomon Four-Group Design

<table>
<thead>
<tr>
<th>Random Assignment</th>
<th>Pretest</th>
<th>Lecture</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group 1</td>
<td>70</td>
<td>Yes</td>
<td>85</td>
</tr>
<tr>
<td>Control group 1</td>
<td>70</td>
<td>No</td>
<td>70</td>
</tr>
<tr>
<td>Experimental group 2</td>
<td>0</td>
<td>Yes</td>
<td>85</td>
</tr>
<tr>
<td>Control group 2</td>
<td>0</td>
<td>No</td>
<td>70</td>
</tr>
</tbody>
</table>

Table 11-2 presents another possibility. Based on the difference between pretest and posttest scores, it shows that the attitudes of the control group also changed. Why? They did not attend the lecture. Is it possible that the pretest influenced their attitudes toward welfare recipients? (For the sake of our example, the control group did not read, view programs, or learn about welfare in any way.) If the answer is yes, can you assume that the change in the scores experienced by the test group was also impacted by the pretest? In other words, did the lecture and pretest interact to generate the change in attitudes? You can use the Solomon Four-Group Design to answer these questions.

Table 11-3 presents another possibility. Note that the posttest scores used to determine the occurrence of a change in attitude for the two experimental groups are the same. This is what you would expect if the pretest did not interact with the lecture stimulus to impact posttest scores. In other words, lecture attendance is the sole cause of attitude change. In addition, the posttest scores for the two control groups are the same. This suggests equivalent groups and, of course, that their attitudes toward welfare recipients were not influenced by a lecture.

Now consider Table 11-4. These results occur because of the pretest, the lecture stimulus, and interaction between the two.

How much of the attitude change is a result of the pretest, the lecture stimulus, and interaction between the two? You can use the following formula to determine the extent of change resulting from these factors.

\[ I = C - (P + T), \]
where
\[ I = \text{interaction effect} \]
\[ C = \text{difference in scores of experimental group 1} \]
\[ P = \text{pretest effect (control 1 posttest result minus control 2 posttest result)} \]
\[ T = \text{treatment effect (experiment 2 posttest result minus control 2 posttest result)} \]

Applying our formula to the results depicted in Table 11-4 we get the following:
\[ I = (85 - 70) - [(75 - 70) + (75 - 70)] = 5 \]

Our calculation reveals that five points of the fifteen-point change in the scores of experimental group I resulted from the interaction between the pretest and the lecture stimulus. One last point: Our formula does not consider the impact of other factors. To keep things simple, we assume that change is a result of the pretest, the lecture, and any interaction between the pretest and lecture.

**The Posttest-Only Control Group Design**

The classic and Solomon Four-Group designs may be strong experimental designs, but their use is limited because of cost and concern about the reactive effect of the pretest. Thus, some researchers use the Posttest-Only Control Group design, which is a variation of the other two. This design requires the random assignment of participants/subjects to the experimental and control groups. It differs from the other two designs because a pretest is not administered. The design is diagramed in Figure 11-3.

Political scientists use the Posttest-Only Control Group Design more than any other type in their research efforts. Not only is it the easiest one to use, but it also can be used to answer a variety of political questions. Some examples will give you an idea about the broad applicability of this design.
You can use the Posttest-Only Control Group Design in a political participation study that shows that an individual’s (test group) characteristics such as education, gender, income level, and so on (independent variables) cause him or her to participate (dependent variable) differently from other individuals (control group).

You can also use the design in a policy analysis study that shows that the policy output (independent variable) recipients (test group) behave (dependent variable) differently from those (control group) who did not receive output from the policy.

The design also has applicability in a study that shows that the attendees (test group) of a lecture about welfare (independent variable) have more knowledge (dependent variable) about welfare than do nonattendees (control group).

The Posttest-Only Control Group Design has several benefits. It allows you to make causal inferences because you treat the test group prior to measurement of the dependent variable. You can also argue that differences between the dependent variables for the two groups resulted from exposure to the independent variables. Remember, you randomly assign individuals to each group to control for the impact of other factors.

The alert student, however, may have some concern about our latter statement. How sure are you that the random assignment of individuals to your groups made them identical before you introduced the independent variable? Could the posttreatment differences between the groups be a result of pretreatment differences? Unfortunately, this design will not tell you how much of the posttreatment difference results from pretreatment contrasts. This is one reason why some political scientists prefer the classic and the Solomon Four-Group designs.

Frankfort-Nachmias and Nachmias (2000) argue, however, that the design will control for several sources of invalidity that we cover in Section 11-7. Testing and instrumentation, for example, are irrelevant concerns because there is no pretest. In addition, concerns about bias due to selection method are controlled by the random assignment of individuals to the groups.

11-5 Nonexperimental Designs

While experimental designs enhance comparison, control, and the manipulation of variables, their use is often limited. An important criterion of experimental designs, for example, is the random assignment of cases to the test and control groups. At times, however, it is not possible to randomly select and assign cases to these groups. When it is not possible to guarantee randomness, many political scientists use a nonexperimental design, or quasi-experimental design. We also use these designs when we cannot incorporate some of the other features of experimental designs. In many designs, for example, it is difficult to control exposure to the experimental stimulus. In addition, political scientists are most often interested in real-life problems that you cannot observe in the confines of a controlled environment. Despite these limitations, however, experimental designs help us understand the logic of all research designs. They are the model we use to evaluate other designs. They also allow us to draw causal inferences about the variables examined in the study. We cannot do this as easily with nonexperimental designs. Thus, if you understand the structure and logic of the experimental designs, you can also understand the limitation of the nonexperimental designs we are about to discuss.
11-5a Nonrandomized Control Group Design

You use this design when you want to investigate a situation where random selection and assignment are not possible. It is one of the strongest and most widely used nonexperimental designs and is sometimes called a contrasted group design. It differs from experimental designs because test and control groups may not be equivalent. In other words, there is no random selection and assignment. You can determine some degree of equivalency between the two groups, however, by comparing the two pretest results. Figure 11-4 is a model of the Nonrandomized Control Group Design. The model is very similar to the Classic Experiment Design depicted in Figure 11-1. If our teacher of the social policy seminar did not divide her groups randomly, she would be using the Nonrandomized Control Group Design.

11-5b Time Series without Control Group Design

At times it will not be possible for you to work with an experimental group and a control group. One way for you to compensate is to take a series of initial observations prior to introducing the independent variable. The initial observations establish a baseline. Once we establish the baseline, we introduce the independent variable and perform several posttest observations. A substantial change in the behavior of the group makes the independent variable suspect as to the cause of the behavioral change. You can acquire more validity by repeating the experiment in different settings under different conditions. As with the Nonrandomized Control Group Design, this design does not randomly assign cases to the group. Figure 11-5 is a model of the Time Series without Control Group Design.

An example using this type of design might be a study dealing with the relationship between a state’s rate of homicide and the use of capital punishment. Several states, for example, used the death penalty as a possible punishment for homicide prior to the Furman v. Georgia U.S. Supreme Court case of 1972. In Furman, the U.S. Supreme Court found that death sentences were being imposed in an arbitrary and unconstitutional manner. Thus, the imposition of the penalty was curtailed across the nation. Four years later, however, the Court, in Gregg v. Georgia, upheld a capital punishment statute that was rewritten in response to Furman. Using a time series design, one could collect homicide data for several years (pretests/baseline) prior to Furman (treatment/independent variable) and homicide data for several years (posttests/dependent variable) after Furman. If there is no significant difference in the two sets of data, one could conclude that the treatment had no effect on the murder rate. On the other hand, if there is a significant difference, the treatment becomes suspect as to the cause of the change in homicide rates.

time series design: Multiple measurements of a dependent variable before and/or after experimental treatment.
This design is similar to the previous design. It differs because you include a parallel set of observations without the introduction of the independent variable. The additional set of observations enhances the validity of the design depicted in Figure 11-5. Like the other nonexperimental models, random assignment is not a part of this design. Figure 11-6 is a depiction of the Time Series with Control Group design.

An example might be tracking students in an advanced learning curriculum treatment. Based on several aptitude tests given at different times (pretests), several states separate students by placing some in the advanced curriculum classes (experimental group) and relegating the others to the normal curriculum classes (control group). School authorities, in addition to local tests, monitor and compare the progress of the students by analyzing a series of state proficiency tests (posttests). If the scores of the advanced children exceed the scores of the other students, school authorities cite the curriculum as the reason.

11-6 Other Types of Research Designs

In addition to the designs we discussed in the previous sections, there are other ways that we can conduct research. Many political scientists, for example, carry out case studies or go out into the “field” to conduct research.

11-6a The Case Study

A case study is an in-depth examination of an event or locale. When you use this type of design you do not assign cases to a test or control group. An example might be an analysis of some public agency where you monitor client satisfaction for a specified period of time. You then consolidate service delivery based on customer recommendations. After some time period, you start to monitor client satisfaction again. You find that their level of satisfaction has significantly increased. What is the reason for the increase? Your steps to consolidate service delivery have to be the reason, of course.

There are some inherent problems with the case study. External validity (see Section 11-7c, External Validity), for example, is poor. It is difficult to generalize...
the results of a study that concentrates on a single unit of analysis. What works for one agency might not work for another one. In addition, it is difficult to control the impact of external factors. Service delivery might improve because of new energetic personnel and not because of the consolidation of service delivery.

11-6b Field Research

When you study the behavior of a group of people in their natural setting you are doing field research. Field research requires you to participate in the lives of those you want to study so that you can better understand their behavior. As the investigator doing field research, you can act as a complete participant or as a participant observer. In the first role, the research objectives are unknown to those you observe. You try to become a member of the group and conceal your identity and purpose from the group members. In the latter case, you become an active member and participant in the group. You also, however, make your presence known to those you want to study.

Many political science field experiments are somewhat different from the type just discussed. Field research for policy analysts is an extension of the classic design into a natural setting. They are not, therefore, a separate type of design. Analysts attempt to control the selection and assignment of subjects to groups and the manipulation of the independent variable.

An example of a field experiment is a state of Texas experiment in job placement training for prisoners about to be released from confinement. The Texas Department of Corrections and Department of Employment, with the assistance of Texas A&M University, conducted a field experiment to test the effects of in-prison vocational placement training on the capabilities of parolees to obtain employment upon release from prison. Texas officials thought that the vocational training would give recipients a head start in the job market and help reduce recidivism.

The research design was relatively simple. Parolees living in Dallas, Houston, and San Antonio made up the pool of possible subjects. Those inmates who received job placement training (the independent variable) made up the experiment group. The control group, those who did not receive the training, consisted of parolees who had similar social and psychological profiles as the members of the experiment group. Researchers also tried to interview members of both groups within seventy-two hours of their release. This was an attempt to control for possible extraneous factors. It also served as a type of pretest. Questions included the number of job contacts made by the parolee, current personal income level, and the income level of anyone else living with the individual. The posttest, conducted six months after the initial interview, was a follow-up interview designed to determine the work history of the parolee since their release. Analysts then compared the work histories of both groups to see if those who received the placement training had a better work history than those who did not receive the training. If there was a significant difference between the two groups, the state of Texas would increase the budget for the job placement program.

There are several problems associated with field experiments. The most obvious deals with the subject’s environment. In a true classic experiment design, you have total control over the subject’s environment. In field research you are not in complete control of the environment. It is difficult to control, account for, and distinguish between the impact of environmental changes on the subject’s behavior and the impact of the experimental stimulus. Can you be sure, for example, that the job placement training was the major reason for a more stable work history? How do you control for the impact of the intervening variables introduced in each parolee’s life during the six months between their release and posttest?
The analysts and sponsors of this particular study wanted to know if their program would benefit the citizens of Texas. Therefore, a study about the behavior of parolees in Dallas, Houston, and San Antonio could have application to the entire state. But can you generalize the results of this study to other parts of the country? It is difficult to generalize the findings of a regional program to a national level. The vocational behavior of a group of parolees in Dallas, Texas, for example, might not represent the vocational behavior of parolees in New York City, Chicago, or San Francisco. In other parts of the country the vocational behavior might be impacted by the state of the economy, public transportation opportunities, and the political viewpoints of government officials.

Participation in this program was voluntary. Prisoners could apply for the in-prison program. There was, however, neither room nor time to give the training to all of those who volunteered. Thus, an important ethical issue arose. Should not a program intended to enhance the assimilation of prior offenders into the mainstream of society be available to all upon request? After all, analysts measured the success of the program by comparing the employment and recidivism histories of the subjects.

11-7 Obstacles to a Successful Research Effort

There are several obstacles you need to consider when carrying out your research efforts. In the following sections, we spend some time discussing internal validity and external validity.

11-7a Internal Validity and Extrinsic Obstacles

**Internal validity** addresses the question of whether the independent variable did, in fact, cause or contribute to changes in the dependent variable. Did the manipulation or variation in the independent variable make a difference in the dependent variable? The criteria for assessing internal validity are time order, covariation, nonspuriousness, and theory.

There are several extraneous and innate obstacles to achieving internal validity (Frankfort-Nachmias and Nachmias 2000, 95–96). Some bias can occur because of the way you select and assign individuals to your test and control groups. These extrinsic factors produce differences between the groups before you apply the experimental stimulus. As a result, you are uncertain about the difference between the groups after you apply the independent variable. Is the difference a result of the stimulus or factors related to the selection and assignment procedures? Therefore, you need to be aware of ways to control selection bias.

**Matching**

The purpose of **matching** is to enhance the possibility that you assign subjects to test and control groups based on characteristics you know are related to the research study. There are two basic ways to match test and control group participants. Pair-wise matching requires you to assign pairs of subjects matched on relevant characteristics to each group. For each person you assign to the test group, you try to assign an individual with identical characteristics to the control group. The researchers involved in the Texas program dealing with job placement training in Texas prisons used this method of matching. As a means of controlling the effect of education, for example, for every parolee in a specific education category in the test group, a parolee in the same category was placed in the control group. Differences noted between the occupational achievements of members of the two

---

**internal validity**: The ability to show that manipulation or variation of the independent variable causes the dependent variable to change.

**matching**: A control method that involves making the experimental and control groups equal in accordance with extrinsic factors that are presumed to relate to the research question (race or age, for example).
The Research Design

groups were attributed to the program and not to level of education, which was one of many extrinsic factors that were controlled.

The pair-wise procedure has a major limitation. Sometimes it is difficult to identify and match participants on important characteristics. For example, the researchers in the Texas penal study wanted to control for age, gender, race, education, and a psychological profile score calculated by the Texas Department of Corrections. Thus for every twenty-five-year-old Anglo male with eight years of education and a profile score of 4 in the test group, researchers had to find an individual with the same combination of characteristics for the control group. Often, it was a monumental task that required researchers to pore through thousands of parolee records. Thus, while pair-wise matching is precise, it is also inefficient and cumbersome.

Frequency distribution matching is another way to assign subjects to test and control groups. Although less precise, it is easier than pair-wise matching. You do not match the groups based on one-to-one matching. Instead, you match the groups based on the average of each extrinsic characteristic. If the average level of education for the test group is ten years, then the average for the control group should be ten years. If you control for race, then each group should have the same proportion of Anglos, Latinos, and African Americans, for example. Table 11-5 shows an example of frequency distribution matching.

Matching raises another question: How do you know you considered all the relevant factors during the selection and assignment stage? By this time we are certain that you know the answer. You must familiarize yourself with the literature in order to discover characteristics others have found important in accounting for differences between individual behaviors.

The bottom line, however, is that matching does have some limitations. Thus, random assignment is a better way to select participants for your study.

**Random Assignment**

Matching helps you control for a limited number of predefined factors that you identified through your literature review. As a result, you think you have identified all of the characteristics others considered important in explaining differences in behavior. But are you sure you identified all of the important ones? No, you are not sure. There may be other variables that you do not know of that could lead to erroneous causal interpretations. So what do you do? Many political researchers resort to randomization to construct their sample. Randomization controls for the influence of characteristics you might or might not know about. It is a process whereby

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Experimental Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Median; 22</td>
<td>Median; 22</td>
</tr>
<tr>
<td>Gender</td>
<td>Male; 49%</td>
<td>Male; 49%</td>
</tr>
<tr>
<td>Race</td>
<td>African American; 12%</td>
<td>African American; 12%</td>
</tr>
<tr>
<td>Education</td>
<td>College; 15%</td>
<td>College; 15%</td>
</tr>
<tr>
<td>Income</td>
<td>$20,000–$30,000; 35%</td>
<td>$20,000–$30,000; 35%</td>
</tr>
<tr>
<td>Religious Affiliation</td>
<td>Catholic; 28%</td>
<td>Catholic; 28%</td>
</tr>
<tr>
<td>Marital Status</td>
<td>Divorced; 15%</td>
<td>Divorced; 15%</td>
</tr>
<tr>
<td>Region</td>
<td>Midwest; 25%</td>
<td>Midwest; 25%</td>
</tr>
</tbody>
</table>

**Purpose:** To have demographically similar groups to address selection and assignment bias.
all cases or subjects in the population have an equal chance of selection and inclusion in the study. It, perhaps, is the foremost method for equating experimental and control groups. We discuss randomization in more detail in Chapter 12.

11-7b Internal Validity and Intrinsic Threats

Researchers have found that, in addition to their experimental stimuli, there are other intrinsic factors that account for changes in their research subjects. These factors can pose a threat to the study’s internal validity. Many researchers have identified the following intrinsic factors that could nullify causal findings.

History

You want to control your research environment so that exposure to the experimental stimulus is the sole explanation for differences between the test and control groups. Over the period of the study, however, other events might occur that also account for change in the dependent variable. History pertains to events that occur throughout the study that might affect your subjects and provide competing explanations for changes in the dependent variable. Did your subjects change their viewpoints about gun control because you exposed them to violent crime statistics? Or did their viewpoints change because they were crime victims during the study period? The longer the period of time between the pretest and posttest, the greater the possibility that external events will taint the study.

Testing

Experimental research requires a comparison of pretest and posttest results to determine the impact of an experimental stimulus. On occasion the pretest could impact the study. A pretest might sensitize your subjects and contribute to better scores on the posttest. As a result, you encounter the problem of testing, which is another threat to the internal validity of your study. What caused a change in the dependent variable, the pretest or application of the independent variable? Consider this example. Your friend took the state bar examination and failed. She then enrolled in a seminar designed to help individuals prepare and take the bar exam. When she retook the exam, she attained a passing score. Why? Was it the seminar? Was it study sessions with friends? Or, did her experience with the first exam prepare her? In any case it is difficult to attribute her success entirely to the seminar. As discussed, if you are concerned about the testing effect you could use the Solomon Four-Group Design or the Posttest-Only Control Group Design.

Regression Artifact

You might assign someone to your test or control group because of a poor score on the pretest. But what if the subject is having a bad day and the score is not indicative of his or her ability? If the individual retook the test, the score might be better. True, you might attribute the increase to testing, but you also might attribute it to the problems the subject experienced at the time of the first test. With both explanations possible, you cannot be sure as to the reason for an enhanced score on a subsequent test. This is an example of a regression artifact, which is another intrinsic factor that can account for changes in research subjects.

Experimental Mortality

When you assign subjects to control and test groups, you try to make the groups equivalent. Over the course of the study, however, some might drop from the
study. As a result the composition of the groups may change enough to bias the results. Experimental mortality is most acute during a study that covers several months.

**Instrumentation**

The problem of instrumentation, or instrument decline, deals with changes in the measuring instrument between the pretest and posttest. To attribute change to the independent variable, you must show that repeated uses of the instrument will produce equivalent results. In other words, you want to make sure the instrument you use to collect data is totally reliable. When applied by different persons, other factors, such as the attitude of the person administering the test, should not influence the results. For example, if a panel of political sociologists used pre-seminar and post-seminar evaluations to analyze the effect a seminar had on the attendees’ level of political tolerance, any changes in a panel members’ basis of evaluation that occurred between evaluations could bias the findings.

**Maturation**

Over time, people change because of psychological, social, or biological processes. These changes could affect your subjects and create differences between the test and control groups. Therefore your groups will no longer be equivalent. This could cause you to make erroneous inferences. Was the increased racial awareness of your subjects attributed to a government program? Or was the increase due to increases in subjects’ ages and increased exposure of subjects to members of other racial groups? Thus, maturation is another problem you must consider when evaluating people’s behavior.

**11-7c External Validity**

External validity is the extent that you can generalize and apply your findings to larger populations and different settings. You want to know whether a study of American college students at a particular university applies to all college students across America, for example.

Several factors often limit the external validity of experimental designs. The artificial setting of the experiment is one limitation. Often conditions that hold in experimental settings will not hold in real-life settings. The experimental setting is too controlled and too pristine.

Occasionally experimental subjects react differently because they realize they are part of an experiment. This could also taint the study results. This possibility occurred during the Hawthorne Electric Plant studies conducted during the late 1920s and early 1930s. A team of researchers from the Harvard Business School wanted to analyze the impact of workplace environmental enhancements on productivity. Their findings showed that the productivity of employees exposed to better conditions did increase. The productivity of employees not subjected to the better conditions, however, also increased. What was the answer? The researchers hypothesized that the novelty of participating in an experiment, receiving attention, and increased cohesiveness of study participants contributed to the increased productivity. In sum, the study participants worked hard to make the study a success. They worked so hard that they tainted the results of the study. They also, however, made researchers take steps to address the complex interaction of the human dynamic with the environment in future studies.
11-8 A Comparison of Research Designs

Table 11-6 shows that experimental designs differ from nonexperimental designs in several ways. The strength of experimental designs is that they meet the requirements of causality. They exert a great deal of control that is manifested through matching and randomization. You can control for the effect of known and unknown extrinsic characteristics or variables. There is also a great deal of control over the introduction of the experimental stimulus. Thus, experimental designs are strong on manipulation. This allows you to establish causal explanations for political behavior more easily than with nonexperimental designs. The classic experiment and Solomon Four-Group designs are also strong on comparison. You can compare the impact of independent variables by looking at changes in behavior between groups or within a single group.

In sum, experimental designs are stronger in internal validity than nonexperimental designs. You have more control over the independent variables, the units of analysis, and the environment in which the behavior occurs.

The strong points of the experimental designs are the flaws of the nonexperimental designs. Without random assignment to groups, control over competing explanations is impaired. Difficulties in establishing the time order of events also detracts from your ability to manipulate the independent variables. In addition, the absence of control groups in some of the designs severely limits your ability to compare results. Consequently, your ability to draw clear inferences from your study is reduced.

The experimental designs, however, also have some limitations. A common complaint is that the results attained from experimental designs are tainted and inappropriate for the real world. It is difficult to reproduce a real-world application in the laboratory environment. Because of the sanitized or laboratory setting and the extent of control you use, it is more difficult to generalize the results to other populations. As we discussed, political scientists use nonexperimental designs because they deal with real-world situations. Consequently, their findings can be generalized. Thus, a strength of nonexperimental designs is their external validity.

Many have also pointed out that experimental designs are weak in representation. That is, many experiments include volunteer subjects and as a result are not representative of the population. This problem makes it difficult for you to generalize the results to the public (Frankfort-Nachmias and Nachmias 2000, 102).

While no design is perfect, you can work to improve the findings of your research project. You can enhance experimental results by defining your population and by using a probability sample to select cases for study. You can enhance nonexperimental results by identifying and controlling as many opposing explanations as possible.
Table 11-6 A Comparison of Research Designs

<table>
<thead>
<tr>
<th>Design</th>
<th>Goal</th>
<th>Model</th>
<th>Compare</th>
<th>Manipulate</th>
<th>Control</th>
<th>Generalize</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic</td>
<td>Study the effect of an influence on a carefully controlled study.</td>
<td>T, R, S, PST.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, R, P, PST.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solomon Four-Group Design</td>
<td>Minimize the Hawthorne and testing effects</td>
<td>T1, R, P, PST.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1, R, P, PST.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2, R, S, PST.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2, R, PST.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest Only</td>
<td>Evaluate a situation that cannot be pretested.</td>
<td>T, R, S, PST.</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, R, PST.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonrandom Group Design</td>
<td>Investigate a situation where random selection and assignment are not possible.</td>
<td>T, P, S, PST.</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, P, PST.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Series Design</td>
<td>Determine the effect of a variable introduced after initial observations with only one group.</td>
<td>T, Pn, S, PSTn.</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>without a control group</td>
<td></td>
<td>T, Pn, S, PSTn.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Series Design</td>
<td>Enhance the validity of the Time Series Design without a control group by using a control group.</td>
<td>T, Pn, S, PSTn.</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>with a control group</td>
<td></td>
<td>C, Pn, PSTn.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key
T = test group.               C = control group.                T1, T2, C1, C2 = test/control groups 1 and 2.
P = pretest.                   PST = posttest.                   
R = random assignment.         n = numerous tests.                 
S = stimulus application.      1 = very low; 2 = low; 3 = moderate; 4 = high; 5 = very high.

Chapter Summary

In this chapter we introduced you to the different types of research designs you can use to complete a research study. Research designs provide you with a plan to use when collecting, analyzing, and interpreting phenomena you believe provide answers to your research problem. As such they are an important part of the composite research process.

We discussed several types of designs. We also highlighted the advantages, disadvantages, and differences between the experimental and nonexperimental designs. Experimental designs are stronger in internal validity but weaker in external validity than the nonexperimental designs. Thus, an experimental design enables you to enhance the comparison, manipulation, and control of your study’s components. On the other hand, you cannot generalize the results as readily as you can with a nonexperimental design.

We also discussed several obstacles to the internal and external validity of your study and presented ways to overcome those obstacles. History, maturation, regression artifact, and testing can taint the internal validity of your study. A clear definition of your population, probability sampling techniques, and a shorter period of study will help overcome these internal validity threats. You must also be mindful of the way you select and assign members to your test and control groups in order to enhance the external validity of your study.
Chapter Quiz

1. Comparisons, one of the four necessary components to show causation is lacing with ________________ design.
   a. preexperimental
   b. classic experimental
   c. Solomon Four-Group
   d. Posttest-Only Control Group

2. The basics of an experimental research design include
   a. manipulation.
   b. control.
   c. comparison.
   d. all of choices a through c.

3. A major purpose(s) of the Solomon Four-Group Design is to
   a. minimize the Hawthorne effect.
   b. minimize pretest influence.
   c. minimize posttest influence.
   d. Both choices a and b are major purposes of the Solomon Four-Group Design.

4. When political scientists study the behavior of a group of people in their natural setting, they may be conducting
   a. a case study.
   b. field research.
   c. a classic experiment.
   d. a time series analysis.
   e. both choice a and choice b.

5. If you are concerned about the testing effect on your research results, you could use the ________________ to address your concern.
   a. Solomon Four-Group Design
   b. classic experimental design
   c. Posttest-Only Control Group
   d. solutions in both choices a and b
   e. solutions in both choices a and c

6. ________________ is the extent that a political scientist can generalize and apply his or her findings to larger populations and different settings.
   a. External validity
   b. Internal validity
   c. External reliability
   d. Internal reliability

7. ________________ pertains to events that occur throughout the study that might affect your subjects and provide competing explanations for changes in the dependent variable.
   a. Maturation
   b. History
   c. Regression artifact
   d. Experiment mortality

8. ________________ addresses the question of whether the independent variable did, in fact, cause or lead to changes in the dependent variable.
   a. External validity
   b. Internal validity
   c. External reliability
   d. Internal reliability

9. The classic experimental research design is strong on ________________ but weak on ________________.
   a. internal validity; external validity
   b. external reliability; internal reliability
   c. internal reliability; external reliability
   d. external validity; internal validity

10. ________________ occurs when one or more of your research subjects “drop out” of the experiment.
    a. Maturation
    b. History
    c. Regression artifact
    d. Experiment mortality

Suggested Readings