Ikon Images/Alamy Stock Photo

8

EXPERIMENTAL DESIGNS

LEARNING OBJECTIVES

Performance Objective

Prepare an experimental design scenario as an alternative proposal for your term project based on the experimental knowledge you have acquired. Focus on problems most commonly encountered in business.

Enabling Objectives

1. List the major classifications of experimental design.
2. Specify the three conditions and four criteria indispensable for causality.
3. Distinguish between internal and external validity and identify threats to each that endanger the results of a study.
4. Explain why a true experimental design is regarded as the most accurate form of experimental research and describe the role of a control (or comparison) group in relation to the treatment.
5. Support the position that it is unethical to withhold treatment from a control group in instances where the treatment provides substantial benefit.
6. Classify other randomized designs that rely on random assignment to produce group equivalence, to balance treatment and control groups, or to compensate for confounding variables.
7. Explain the advantages of a quasi-experiment and how it promotes increased realism and ecological validity when conditions only vary naturally (not by researcher manipulation).
8. Illustrate how matching and other techniques balance treatment and control groups to reduce the effect of confounding variables.

In this last chapter on design, I introduce many of the features of the *experimental model* and also discuss variations in experimental design. This chapter also covers causality, validity, and the use of matching and other mechanisms to balance treatment and control groups when random assignment is not possible.

OPTIMIZING BUSINESS EXPERIMENTS

In a recent *Harvard Business Review* (*HBR*), Thomke and Manzi discussed business experiments they had conducted or studied during their 40-plus years of collective experience with companies. Those firms included Bank of America, BMW, Hilton, Kraft, Petco, Staples, Subway, and Walmart.[1] Thomke and Manzi's advice was considered so valuable for their readers that *HBR* elevated the article to their "10 Must Reads"

and it was also a McKinsey Awards Finalist. The authors suggested five questions that companies should answer before beginning a *business experiment*. Their guidance should entice you to consider experimental design or at least be knowledgeable enough to recommend it as a rigorous test to determine if a new product or program will succeed. Here is a quick summary of their suggestions and a few examples from their article.

1. *Does the experiment have a clear purpose?*

When executives disagree on a proposed action, an experiment may be the most pragmatic way to answer the question, provided that the hypothesis is stated in unequivocal terms. (We addressed the importance of specificity in research questions and hypotheses in Chapters 2 and 3.) In 2013, Kohl's did just that by testing the following hypothesis: "Opening stores an hour later to reduce operating costs will not lead to a significant drop in sales." Results of the experiment involving 100 Kohl's stores supported the hypothesis.

2. *Have stakeholders committed to abide by the results?*

This requirement reduces dissension on the management committee and prevents influential persons from selecting evidence that only supports their point of view. Publix Super Markets, which does business predominantly in the Southeastern United States, has a procedure to ensure commitment. A proposal is submitted for financial analysis and then to a committee, which includes the finance executive. Finance then approves programs that have followed the process and have positive experimental results. Other companies with similar approval protocols carefully evaluate the cost-benefit of testing.

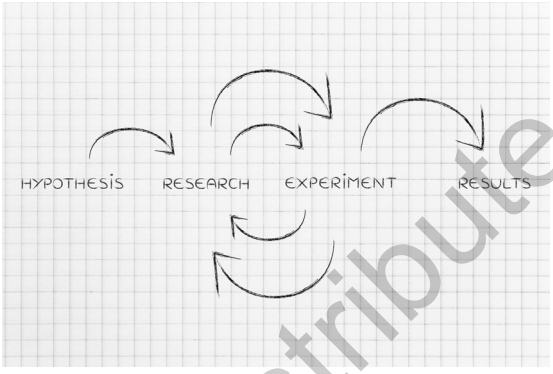
3. *Is the experiment doable?*

The business environment's complexity (from the supply chain to the distribution channels to end users) makes it challenging to sort out the cause-and-effect relationships of the variables under investigation, especially because the environment changes rapidly and confounding variables, not just the IV and DV, thus change as well. Large samples address this problem regarding effect size but are often impractical because of scaling the experiment's size and cost.

4. *How can we ensure reliable results?*

Thomke and Manzi suggest three ways to increase validity and reliability: randomized field trials, blind tests, and big data. With *randomization* in the field, companies may take a large group of individuals with the same characteristics and randomly divide them into test and control groups. Capital One, for example, has a history of demanding field trials even for questions like the color of envelopes for product offers. Petco previously selected

its 30 best stores for treatment and the 30 worst as controls. As you might expect, the results were impressive but failed at launch. Now Petco, along with Publix and others, includes customer demographics, competitor proximity, and store size to get more valid and reliable results. *Blind tests* reduce participant perceptions that deviations are occurring, thus causing them to behave differently. Often, a company's employees are not aware of an ongoing experiment. (We define single, double, and triple blinding later in the section on internal validity.) *Big data* are useful in resolving disputed results. Take a company in which different groups produced conflicting results in separate experiments on the same program. According to Thomke and Manzi, "To determine which results to trust, the company employed big data, including transaction-level data (store items, the times of day when the sale occurred, prices), store attributes, and data on the environments around the stores (competition, demographics, weather)."[2]



AP Photo/Fotolia

Norwegian economist Trygve Haavelmo, who won the 1989 Nobel Prize, observed that there are two types of experiments: "those we should like to make" and "the stream of experiments that nature is steadily turning out from her own enormous laboratory, and which we merely watch as passive observers." If firms can recognize when natural experiments occur, they can learn from them at little or no additional expense. For example, when an apparel retailer opened its first store in a state, it was required by law to start charging sales tax on online and catalog orders shipped to that state, whereas previously those purchases had been tax free. This provided an opportunity to discover how sales taxes affected online and catalog demand.

Source: Anderson, Eric T., and Duncan Simester, "A Step-by-Step Guide to Smart Business Experiments," *Harvard Business Review*, March 2011, https://hbr.org/2011/03/a-step-by-step-guide-to-smart-business-experiments.

5. Have we gotten the most value out of the experiment?

Because of the diversity of customers, markets, and geographies for many retail companies, the "where" question is critical. For example, Petco's initiatives are used in stores that are the most similar to the test stores that produced the best results. The issue of exploiting the captured data is also important. Previously, Publix had an 80:20 ratio of testing time versus analysis. Their goal is to reverse that ratio, thereby extracting more useful information.

The authors conclude by answering why experiments are valuable for business decision-making:

> The lesson is not merely that business experimentation can lead to better ways of doing things. It can also give companies the confidence to overturn wrongheaded conventional wisdom and the faulty business intuition that even seasoned executives can display. And smarter decision-making ultimately leads to improved performance.[3]

Do not neglect considering an experimental design for a business study. There is a rich history of experimental design in marketing, such as consumer behavior, advertising, retail store environments, sales, and partner satisfaction in marketing alliances. Experiments are also used in many subfields of management as well as in economics/international economics where studies include individual choice, game theory, and the organization and functioning of markets.

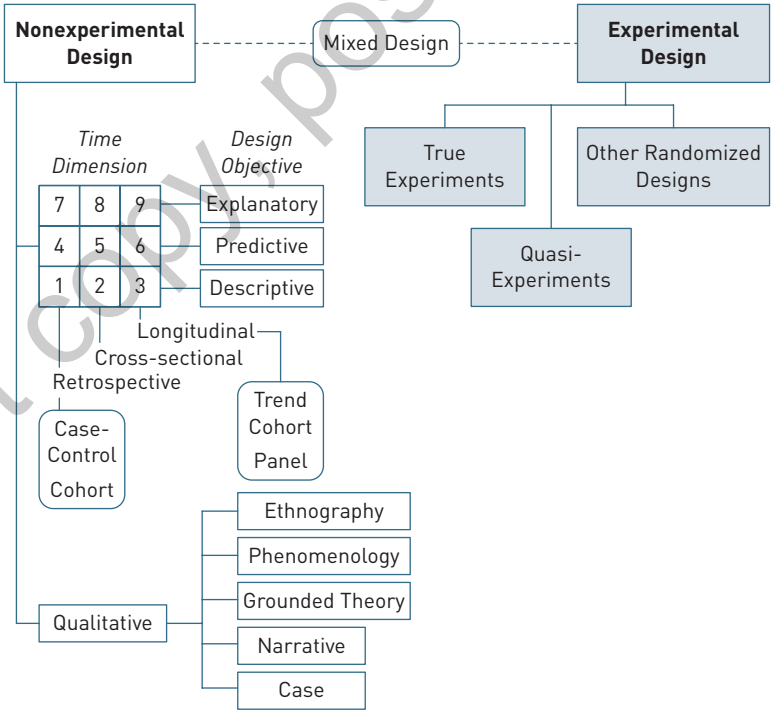
EXPERIMENTAL CLASSIFICATIONS

Campbell and Stanley observed, "By experiment we refer to that portion of research in which variables are manipulated and their effects upon other variables observed."[4] *Experimental designs* consist of *true experiments* (sometimes called randomized experiments or randomized controlled trials), *other randomized designs*, and *quasi-experiments*. *Pre-experiments* are discussed only briefly for comparison purposes.

The design of a **true experiment** is a detailed strategy that is planned to reveal cause-and-effect relationships among variables through manipulation, control, and random assignment to groups. Because of its powerful nature in identifying such relationships, this design is often considered the "gold standard" for evaluating other designs. However, hypotheses claiming causal relationships are bold and susceptible to alternative explanations.

The second group of experimental designs, *other randomized designs* (shown in Exhibit 8.1), provide flexibility in handling numerous variables simultaneously and can be applied to a

EXHIBIT 8.1 ■ Research Designs: Experimental



Note: Experimental Design components are highlighted. The shaded boxes represent the coverage of this chapter, whereas the white boxes are design topics from Chapters 6 and 7.

wide range of research questions that involve field settings. These designs result from random assignment of participants to treatment groups or are based on randomization.

The third group is *quasi-experiments*, which eliminate the problem with directionality but participants are *not* randomly assigned and confounding variables that affect participant selection are not removed. Researchers use these designs for their convenience and their relatively less conspicuous and disruptive nature to participants.

In this chapter, I isolate the right side of Exhibit 8.1 due to the sheer quantity of information on experimental design. Because the nonexperimental category includes many quantitative and qualitative designs, the left side of the graphic was explained in Chapters 6 and 7. After you read this chapter, you will have an extensive choice of experimental and nonexperimental designs from which to create your study.

CHARACTERISTICS OF EXPERIMENTAL DESIGN

Upon reading this section, one thing should be evident: "Good design is obvious. Great design is transparent."[5] Let's begin the technical discussion of experimental design by defining the necessary conditions for experiments to make claims of causality: (a) manipulation, (b) control, and (c) random assignment.

Conditions for Claims of Causality

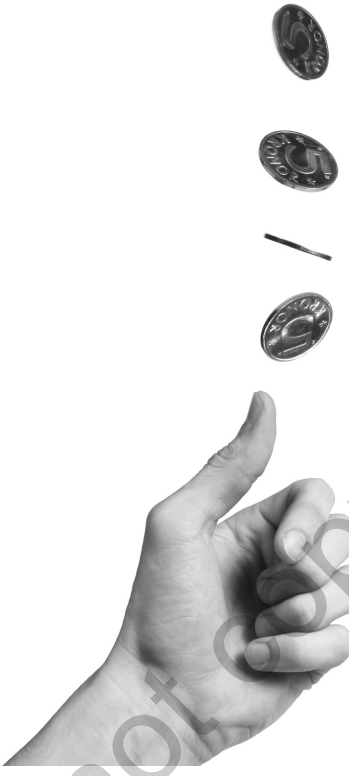
With experimental design, a "defining characteristic is active manipulation of an independent variable (i.e., it is only in experimental research that 'manipulation' is present)."[6] This depiction of **manipulation** suggests that a researcher manipulates or systematically varies the levels of an independent variable (IV) and then measures the outcome of interest, the dependent variable (DV). The manipulated condition (IV) is also known as the "treatment" or "intervention." Levels of the condition are often referred to in shorthand by researchers (e.g., in an experiment of training effectiveness, the three levels are Seminar, OJT, and None). Simplicity and common sense determine the levels of an independent variable. If salary is hypothesized to influence employees exercising stock options, the salary variable might be divided into high, medium, and low, representing three levels of the independent variable. Manipulating an independent variable means changing "its level systematically so that different groups of participants are exposed to different levels

©iStockphoto.com/filadendron

Researchers manipulate or systematically vary the levels of an independent variable (IV) and then measure the outcome of interest, the dependent variable (DV). Here engineer-participants are testing hologram-themed augmented reality glasses where wearers interact with screens and full-color virtual objects. Early models show the repair of light switches with the help of virtual assistants, who draw diagrams and arrows within the picture that the Microsoft HoloLens is seeing.

of that variable, or the same group of participants is exposed to different levels at different times."[7] The term systematic, as used in this statement, implies the existence of procedures to minimize error and bias while increasing confidence in the efficacy of the manipulated treatment. Manipulation is the first feature of the experimental design.

> Notice that the manipulation of an independent variable must involve the *active* intervention of the researcher. Comparing groups of people who differ on the independent variable before the study begins is not the same as manipulating that variable.[8]



Shutterstock/Ronstik

A defining feature of experimental design is random assignment or the assignment of participants to groups (experimental versus control or to different treatment conditions) using a random procedure such as a coin toss with the expectation of getting a 50:50 chance result. Coin tossing is not as useful for randomization as you may think. In a study at the University of British Columbia, 13 participants tossed a coin 300 times trying to achieve a "heads" result. Each participant attained more heads than tails and this difference was statistically significant for seven participants. One of those achieved 68% success with heads.

Note: See Clark, Matthew P.A., and Brian D. Westerberg, "How Random Is the Toss of a Coin?," *Canadian Medical Association Journal* 181, no. 12 (2009): E306–E308.

The second feature is researcher **control** of variables. "In such a design, the researcher considers many possible factors that might cause or influence a particular condition or phenomenon … [and] then attempts to control for all influential factors *except* those whose possible effects are the focus of the investigation."[9] Variables other than the IV and DV are extraneous to the study and presumed to be controlled by randomizing them across participants so that the groups are equal. This effort prevents outside factors from influencing the outcome; however, extraneous variables can creep in at every stage of the process. To be vigilant, researchers can "control extraneous variables through the experimental setting, consent, instructions, sampling techniques, assignment techniques, observation techniques, measurement techniques, interactions with participants, and the use of research designs with control groups."[10]

The third feature of an experimental design is **random assignment** or the assignment of participants to groups (or different treatment conditions) using a random procedure such as a coin toss or random number generator. Random assignment assures that (1) each participant has an equal chance of being assigned to a group and (2) the assignment of one participant is independent of the assignment of another. (*Participant*, in case you missed it in previous references, is the current term referring to an individual taking part in a research study, but I occasionally use the 100-year-old term, "subject," as a time-honored convention.)

Random assignment should not be confused with random sampling, which is an entirely different procedure.[11] Unlike random assignment to groups, which occurs before the experimental condition, **random sampling** from a population aims to ensure that each element in the sampling frame has an equal chance of being included in the sample. When researchers use random sampling, especially in nonexperimental studies, control is enhanced through improved internal validity (by reduction of systematic and random error) and external validity is expanded. "When random samples are not practically possible (you then have a sample in search of a population)."[12] In the most robust experimental designs, random assignment is intended to produce equivalent groups, which is different from a known chance of selection. Random assignment ensures group similarity as the study begins.

Simple Causal Relationships

How do we recognize a causal relationship? David Hume and John Stuart Mill first proposed the criteria that we still use today. Mill's Method of Agreement states that "When two or more cases of a given phenomenon have one and only one condition in common, then that condition may be regarded as the cause (or effect) of the phenomenon."[13] Building on that statement, we find the following four essential **criteria for causality**[14]:

1. The cause (IV) and effect (DV) are related. The first criterion means there needs to be a way to follow the effect back to the cause. If in a factory study, plant layout was not a factor in the production process, then we can't argue that the production layout caused late deliveries to the customer.
2. The cause precedes the effect. A time order must be observed: changes in the IV must happen before changes in the DV. Causal precedence is the temporal antecedence condition.
3. The cause and effect occur together consistently. That is, cause and effect should go together or *covary*. Let's say we have a list of possible subcauses for late deliveries to the customer. These become hypotheses that we plan to test, and they include lengthy preparation time, poorly optimized plant layout, errors in the production process, poor planning, inadequate assembler training, low-quality raw materials, improper maintenance of equipment, and packaging and shipping.[15] When we test low-quality raw materials, we find that late deliveries occur. And if the raw

©iStockphoto.com/vm

A plant's layout and the design of the production process are vital to business operations. An optimized layout boosts production, meets employees' needs, and ensures a smooth workflow. It also optimizes material, machinery, and information flow through a system. On the other hand, a bad layout increases the cost of manufacturing by causing unnecessary handling of materials and movement of equipment and workers.

materials are optimum, then late deliveries are absent. If the cause is inconsistent in its effectiveness, then we should find stronger or weaker effects accordingly.

4. Alternative explanations can be ruled out. The relationship between the IV and DV must *not* be due to a confounding extraneous "third" variable (i.e., no credible third variable accounts for the relationship between the IV and DV, or can cause both).

Let's tie these criteria together with an illustration. Suppose I hypothesize that poor training of assemblers in manufacturing causes late deliveries to the customer. The treatment is specialized training for assemblers by process engineers. Assemblers are randomly assigned to treatment and **control groups** or **comparison groups**. The control group does not receive specialized training (the cause is absent). However, both groups are on identical assembly lines. A pretest-posttest shows that the experimental group's skills have improved. The comparison group shows no effect. If my hypothesis were correct, I would expect this to lower late deliveries. The cause and effect, training and late deliveries, are in proximity; they happen close together in time, so we suspect that they are connected.

However, there is more to it than that. The cause (assembler training) needs to happen before the effect (the decrease in late deliveries). One can demonstrate this by controlling the presence of the cause (training). The cause and effect should occur together consistently. That is, more intensive training should dependably correspond with fewer late deliveries up to a threshold where other factors are responsible for deliveries. We should also be suspicious of the explanation and test our other subcauses (hypotheses) regarding late deliveries to rule out *third variable effects*, such as material preparation time, plant layout, work culture, process errors, poor planning, low-quality materials, and equipment maintenance.[16] We should also be careful with random assignment to experimental and comparison groups as well as confirm identical processes on the assembly lines.

Validity in Experimentation

You judge experimental designs by how well they meet the tests of validity. A design's validity is evaluated by the extent to which it is jeopardized by hazards—or what the experimental literature calls "threats." Campbell and Stanley[17] initially labeled and explained eight threats to internal validity and four threats to external validity. Over the years, the number of threats proliferated, with Cook and Campbell[18] expanding their list to 33 and Shadish et al. settling on 37.[19]

A thorough elaboration of many items is well beyond the coverage of this guide. Thus, I confine my list to four general types of validity and provide explanations for two.

1. **Statistical Conclusion Validity**: the validity of inferences about the correlation (covariation) between treatment and outcome.

2. **Internal Validity**: the validity of inferences about whether observed covariation between *A* (the presumed treatment) and *B* (the presumed

outcome) reflects a causal relationship from *A* to *B*, as those variables were manipulated or measured.

3. **Construct Validity**: the validity of inferences made from the operations in a study to the theoretical constructs those operations are intended to represent. (See Chapter 2 on the nature of "constructs.")

4. **External Validity**: the validity of inferences about whether the cause-effect relationship generalizes to persons, settings, treatment variables, and measurement variables.[20]

Campbell and Stanley advise us that the importance of threats should not be underestimated:

> . . . [T]he first line of attack toward good causal inference is to design studies that reduce the number of plausible rival hypotheses available to account for the data. The fewer such plausible rival hypotheses remaining, the greater the degree of "confirmation." Assessing remaining threats to validity after a study is completed is the second line of attack, which is harder to do convincingly but is often the only choice when better designs cannot be used or when criticizing completed studies.[21]

Internal Validity

An experiment has high internal validity if you have confidence that the treatment has been the source of change in the dependent variable. Internal validity asks, "Do the conclusions we draw about a demonstrated relationship correctly imply cause?" Variables other than the treatment (IV) that influence internal validity are as follows:

- *Extraneous variables* may compete with the IV in explaining the outcome of a study in a cause-and-effect context.
- A **confounding variable** is an extraneous variable that does indeed influence the dependent variable. It can systematically vary or influence the IV and the DV.[22]

Exhibit 8.2 presents 12 primary vulnerabilities to internal validity; more are shown in this chapter's references.

External Validity

External validity is high when the results of an experiment are believed to apply to some larger population. External validity asks, "Do observed causal relationships generalize across persons, settings, treatment variables, measurement variables, and times?"

EXHIBIT 8.2 ■ Sources and Threats to Internal Validity

Threat	Source	Description	Remedy
Maturation	Participants	An alternative explanation is caused by subjects' state of mind, natural change, or development over time: short-term and long-term scenarios.	Control group
Selection		Groups lack equivalency. There are systematic differences in participants other than the presumed cause.	Control group Randomization
Maturation by Selection		The rate of change in specific groups is different over the course of the experiment.	Randomization
Low Construct Validity	Instruments	A measure's construct validity should correspond to an empirically grounded theory and correlate with a known measure possessing *convergent* and *discriminant* validity. You discover a construct's uniqueness through statistical tools like factor analysis.	Reanalyze the instrument
Instrumentation		The instrument, data collection, or observer changes over the course of the study due to unforeseen or careless procedures. Instrument decay (e.g., a physical device's calibration over time) may also occur.	Consistent protocols
Score Regression Toward the Mean		Participants with an extreme score on one test get a lower score on the other test—there is a tendency for scores to move toward the mean. A participant with a high score on the pretest receives a lower score on the posttest and vice versa. Calculating the extent of regression is possible.[a]	Evaluate participants from the sample likely to have extreme scores on the DV
Testing		Participants are sensitized to the IV (e.g., pretesting). Learning effects are not attributable to the IV.	Control group Specialized designs
Experimenter Expectancy	Artifacts	Changes in conscious or unconscious researcher behavior affect participants' response to the IV.	Blind designs[c]
Demand Characteristics		There are differential responses to cues in the experimental and control groups; participants are knowing or expectant about what is happening or is expected to occur.	Cover stories Double-blind designs[c]

Threat	Source	Description	Remedy
Temporal Precedence	**Design/Procedure**	There is ambiguity as to whether the cause precedes the effect.	Accurate temporal manipulation
History		An unforeseen event occurs before or between pretests and posttests of the study (e.g., a union meeting where participants in the study are present and discuss an educational campaign for employees).	Individual testing to partial out the effect (not always possible or efficient)
Attrition (mortality)		Dropout rates are particularly likely in the experimental group and also in the control.[b]	Randomization Retesting

Source: Compiled with input from Campbell, Donald T., and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research*, reprint (New York: Houghton Mifflin, 1963), 8; Leedy, Paul D., and Jeanne E. Ormrod, *Practical Research: Planning and Design*, 11th ed. (Boston: Pearson, 2016), 181, Figure 7.1; and Kirk, Roger E., *Experimental Design*, 4th ed. (Thousand Oaks: SAGE, 2013), 16–21.

Notes:

[a]Prm = 100(1 &– r), where Prm is the percent of regression to the mean and r is the correlation between the two measures. Perfectly correlated variables have no regression effects

[b]See "Attrition" in Exhibit 6.4 from Cook, Thomas D., and Donald T. Campbell, "The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings," Handbook of Industrial and Organizational Psychology, ed. Marvin D. Dunnette (Chicago: Rand McNally, 1976), 223. Experimental attrition occurs because participants are not willing to continue, are not available, have relocated, or are disturbed by the treatment—which sometimes occurs in the experimental group. However, *resentful demoralization* of the disadvantaged occurs when the treatment is desirable, the experiment is obtrusive, and *control group* members become resentful that they are deprived of the treatment and lower their cooperation or leave. Other factors affect equalizing experimental and control groups include the following: (1) *diffusion or imitation of treatment*: if people in the experimental and control groups talk, then those in the control group may learn of the treatment, eliminating the difference between the groups; (2) *compensatory equalization*: where the experimental treatment is much more desirable, there may be an administrative reluctance to deprive the control group members and compensatory actions for the control groups may confound the experiment; and (3) *compensatory rivalry*: this may occur when members of the control group know they are in the control group, which may generate competitive pressures and cause the control group members to try harder.

[c]Blinding occurs when one or more persons are unaware of the intervention. Single blinding refers to blinding of participants *or* investigators. Double blinding refers to blinding of both participants *and* investigators. Triple blinding refers to participants, investigators, *and* data analysts and may also include study writers.

The following are some examples of external validity in checklist form:

- *Population Validity: generalizing* to and across populations
- *Ecological Validity: generalizing* across settings
- *Temporal Validity: generalizing* across time
- *Treatment Variation Validity: generalizing* across variations of the treatment
- *Outcome Validity: generalizing* across related dependent variables[23]

"As a general rule, studies are higher in external validity when the participants and the situation studied are similar to those that the researchers want to generalize to," says Price.[24] There are four potential threats to external validity:

1. *Testing Reactivity*: the interaction effect of testing in which a pretest might increase or decrease participant sensitivity to the IV, making the results unrepresentative for the unpretested population
2. *Interaction Effects*: biases resulting from selection (lack of group equivalency) that interact with the IV (Exhibit 8.2)
3. *Reactive Arrangements*: the effects of people being exposed to the IV in nonexperimental settings
4. *Multiple Interferences*: when participants experience multiple treatments, the effects of prior treatments are not erasable[25]

Researchers strive for a balance between internal and external validity. Too little control reduces their ability to derive causal conclusions; too much control restricts capacity to generalize the results.

PRE-EXPERIMENTS

Although Campbell and Stanley used pre-experiments as a reference point, the pre-experiment does not meet the standards of a bona fide experiment. While it may be useful as an exploratory tool, the pre-experiment is evaluated negatively for threats to internal and external validity. One authoritative source states that "… such studies have such a total absence of control as to be of almost no scientific value."[26]

Accepting this argument, I do not include pre-experiments as designs. They are unique only for comparison with other experimental designs. **Pre-experiments** typically have no control group available for contrast (or an equivalent nontreatment group). The three primitive pre-experiments are as follows:

1. *One-Shot Study:* From this study, it is difficult to draw conclusions because one cannot prove there is a cause-and-effect relationship between the intervention and outcome.
2. *One Group Pretest-Posttest:* This design shows some improvement because a change occurred, but you don't know why because it does not account for an event, maturation, or altered collection method that could occur between data points.

3. *Static Group Comparison:* This is best of the three because it shows that change occurred but is still problematic in the elimination of a control group; groups are *not* equivalent at the beginning (participant selection could result in groups that differ on relevant variables), and it is hard to conclude the reason for observed differences.

TRUE EXPERIMENTS

True experiments (or randomized experiments) are the strongest designs for determining a cause-and-effect relationship. They maximize internal validity. They are also known as *randomized controlled trials* (RCTs) because researchers can control the number and types of intervention. They are a causal study's best defense against counterclaims of alternative causes. During an RCT, the only expected difference between the experimental and control groups is the outcome variable under study. Three essential ingredients of a true experiment were previously described as (a) investigator *manipulation* of the IV; (b) *control* of the study situation, protocol, and setting (including the use of a control group); and (c) *random assignment*. Furthermore, a true experiment should be a study of only one population.

Example: Pretest-Posttest Control Group

To establish a frame of reference, let's look at an example of a true experiment (the pretest-posttest control group design in Exhibit 8.3), as described by Campbell and Stanley.[27] Participants are randomly assigned to experimental and control groups, thereby making the two groups similar. The experimental group is composed of participants receiving the experimental treatment.

EXHIBIT 8.3 ■ Pretest-Posttest Control Group Diagram

Time	t_1	t_2	t_3
Randomize	Pretest	Treatment	Posttest
R_E	O_1	X_E	O_2
R_C	O_3		O_4
Ideal Case	E / C		

Time t_1, t_2, t_3

R random assignment of participants
O the test or observation

E the experimental group
C the control or comparison group

Source: The uniform code and graphic presentation are adapted from Levy, Yair, and Timothy J. Ellis, "A Guide for Novice Researchers on Experimental and Quasi-Experimental Studies in Information Systems Research," *Interdisciplinary Journal of Information, Knowledge, and Management* 6 (2011): 154.

It is possible to have more than one experimental group but that requires using a different design than the one illustrated. True experiments have a control group(s). Control participants are also randomly assigned and created in the same manner as the experimental group, but they do not receive the treatment. The control group provides a reliable baseline for comparison of the treatment's effect on the experimental outcome. How important is this comparison? "Well causality is even more plausible if you can compare [it] to a situation where the cause is absent, showing that the effect does not occur when the cause is absent."[28]

Example: Solomon Four-Group Design

Another notable true experimental design is the Solomon four-group design, shown in Exhibit 8.4. This design has a better reputation because it represents the first overt attempt to address external validity issues.

With E-B and C-2 lacking a pretest, you may determine the effects of testing and the interaction of testing and the treatment (X). Not only is generalizability increased but you may also compare the efficacy of X through four evaluations: $O_2 > O_1$, $O_2 > O_4$, $O_5 > O_6$, and $O_5 > O_3$.[29] The effect of randomization may also be confirmed with an O_1-O_3 comparison.

The other true experimental design is the posttest-only control group design, in which the pretest is said to be nonessential if you subscribe to the notion that the lack of initial bias between groups is a function of randomization.

EXHIBIT 8.4 ■ Solomon Four-Group Diagram

Time	t_1		t_2	t_3
Randomize	Pretest		Treatment	Posttest
R_E	O_1		X_{E-A}	O_2
R_{C1}	O_3			O_4
R_E			X_{E-B}	O_5
R_{C2}				O_6
Ideal Case	E-A E-B			
	C-1 C-2			

Time t_1, t_2, t_3
E_A Exp. Gp. A
E_B Exp. Gp. B
C_1 Control 1
C_2 Control 2

Control Groups

Control groups receive no treatment—the experimental stimulus is withheld, or some standard treatment is used. Control members are selected by random assignment (sometimes matching) to have the same characteristics as the treatment group. Both groups experience

identical conditions during the experiment except the control group is not exposed to the treatment condition. Frankfort-Nachmias et al. assert that an advantage of control groups is that they can reduce threats to internal validity (refer to Exhibit 8.2) as follows:

- History does not become a rival hypothesis because the control and experimental groups are both exposed to the same events.
- Maturation is counteracted because both groups undergo the same changes.
- Instrument change can be prevented with a control group; if the instrument's unreliability produces differences between posttest and pretest scores, this will be revealed in both groups. This solution to instrumentation is only effective when both groups are exposed to identical testing conditions.
- Regarding testing, if the reactive effect of measurement is present, it is manifested in both groups.[30]

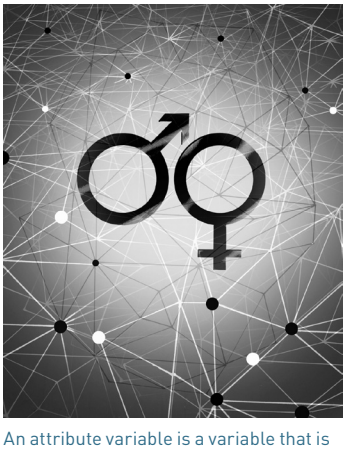
The authors also contend that control groups do not address the issue of attrition (mortality), since one group may lose more participants than the other. However, a control group can help counteract the factors that interact with selection.[31]

True Experiments and Attribute Variables

A true experiment should *not* be used to answer a research question that is not amenable to its requirements. In a series of ongoing studies about women in management, McKinsey & Company concluded with a caveat:

> Companies with a higher proportion of women in their management committees are also the companies that have the best performance. While these studies do not demonstrate a causal link, they do, however, give us a factual snapshot that can only argue in favour of greater gender diversity … [based on the performance factors of return on equity, stock price growth, and operating result].[32]

What if we took this conclusion as a hypothesis and tested it using a true experiment? This is not possible. True experiments require random assignment of participants to different groups. Gender, like other personal characteristics (ethnicity, personality, education, intelligence), is a **subject or attribute variable** and is not changeable. Participants recruited for experiments come to the experimental setting with characteristics established by heredity and environment. Someone cannot be randomly assigned to be a male or female—thus, no



Ikon Images/Alamy Stock Photo

An attribute variable is a variable that is a characteristic or trait of a participant, which researchers cannot manipulate but can only measure. It might be a variable that is fixed like gender, race, or psychological condition. Researchers cannot manipulate any characteristic that is inherent or preprogrammed.

random assignment or manipulation by an experimenter can take place.[33] In contrast, an experiment can help researchers investigate "how participants react to people who vary in these characteristics."

OTHER RANDOMIZED DESIGNS

Adjacent to "True Experiments" in Exhibit 8.1 is a category labeled "Other Randomized-Based Designs." Other **randomized-based designs** rely on random assignment to produce group equivalence, to balance treatment and control groups, or to compensate for confounding variables. Randomization is important to understanding these designs. However, I leave extensive diagramming and coverage of each design to books devoted to that purpose. The previously cited work by Montgomery (*Design and Analysis of Experiments*) provides thorough coverage on designs such as *randomized block*, *Latin square*, *factorial/fractional factorial*, *split plot*, *repeated measures*, *hierarchical*, and *covariance.* You will discover a few others there too.

The types of randomization used to empower these designs include simple randomization (discussed in the section "Conditions for Claims of Causality"), block randomization, stratified randomization, and covariate adaptive randomization.[34]

Simple randomization involves the assignment of participants to control and treatment groups through conventional methods (dice, coin toss, odd-even numbers in a card deck, or random number tables and generators). **Block randomization** results in groups of equal size. Researchers determine the optimal block size for the experiment (sometimes because of the group or cell sizes required by a statistic) and then randomly chose blocks to establish participant assignment. Additional steps are taken to control covariates and restore balance.

Stratified randomization focuses on controlling confounding variables (covariates) that influence the study's outcome. Researchers identify specific covariates through the literature, experience, or foreknowledge of the recruitment pool before group assignment. For example, in a study of hiring decisions, a resume will tell us the length of time a person held a position and the time between jobs. Stratified randomization is "achieved by generating a separate block for each combination of covariates, and subjects are assigned to the appropriate block of covariates."[35]

Covariate adaptive randomization is an alternative to stratified randomization. It assigns new participants to treatment groups involving the technique of "minimization." The process of allocation depends on characteristics of previously recruited participants already assigned. For example, assume you had 20 participants already recruited and the 21st is a male with a high score on a covariate (age). After you compare the treatment and control groups by examining totals for each category, the participant is assigned to the group that produces the most balance. This approach is a form of dynamic allocation or, in our case, "covariate adaptive randomization" because "unlike stratified randomization,

minimization works toward minimizing the total imbalance for *all factors together* instead of considering mutually exclusive subgroups."[36]

Signal Versus Noise and Randomization-Based Designs

Experimental designs may be **signal enhancers or noise reducers**.[37] The signal is analogous to the study variable—the program or treatment being implemented. Noise introduces variability from all the extraneous variables that confuse the strength of the signal. This includes the following: (1) *demand effects* or clues from the setting of the experiment, the equipment, or distractions from the locality; (2) *researcher effects* or cues like nonverbal behaviors of the investigators or impressions that the participants receive about how they should respond; (3) *participant effects* such as a subject's disposition on a particular day, a health irregularity, or prior knowledge possessed by an individual participant, and so forth; and (4) *situational effects* or environmental factors such as temperature, lighting, or discomfort issues. Using a signal-noise ratio, dividing the signal by the noise, the signal should be high relative to the noise. A strong signal (a potent treatment) and accurate measurement (low noise) provide greater likelihood of observing the effect of the treatment or IV. A strong treatment with weak measurement or a weak treatment with strong measurement reduces the effectiveness of the experiment.

Factorial Designs

Factorial designs are *signal-enhancing experimental designs*. A **factorial design** is a randomized experiment (completely or by blocking) using multiple factors to determine their influence on the study's objective. A **factor** is a controlled IV subdivided into levels that are set by the researcher. Each factor must have two or more **factor levels** or values, otherwise it does not "vary." If the factor is tire wear, three levels of the factor might be regular highway, all-weather, and high-performance tires. The levels can cover the full range of offerings or brands or, as in the tire example, just a subset.

Let's consider an example of personality tests, which are increasingly used by human resource professionals. Approximately 2.5 million people take the Myers-Briggs test every year and it is used by 88% of Fortune 500 companies, despite its reliability.[38] Personality questionnaires evaluate how you like to work, relate to others, deal with emotions, and feel about your self-image. Even senior executive candidates at well-known companies take pre-employment tests.

Continuing with the example we just described, suppose we are studying the effect of room temperature in the testing facility and personality test taking.[39] We compare test scores of two independent groups who took the test in an 85-degree room versus those who took it in a 70-degree room. Our experiment has one IV of temperature (not two), but it does have two levels: 70 and 85 degrees. We are also interested in test difficulty, so we add a second IV that also has two levels: simple and complex test difficulty. We now have a 2×2

EXHIBIT 8.5 ■ Factorial Table

	IV-1 Room Temp	
IV-2 Test Difficulty	70 Degrees	85 Degrees
Complex	Complex-70	Complex-85
Simple	Simple-70	Simple-85

between-subjects factorial design. We then randomly assign subjects to four groups, and our matrix is illustrated in Exhibit 8.5.

From the comparison in Exhibit 8.6, you can conclude that the applicants perform better in higher temperatures regardless of test difficulty and they perform better on the simple test regardless of temperature.

In a different scenario, the *shaded* squares in Exhibit 8.7 show an "interaction effect"—crossing lines. Those taking the more complex test performed better under lower temperatures but did worse in the 85-degree condition than those taking the simple version of the test.

EXHIBIT 8.6 ■ Factorial Plot

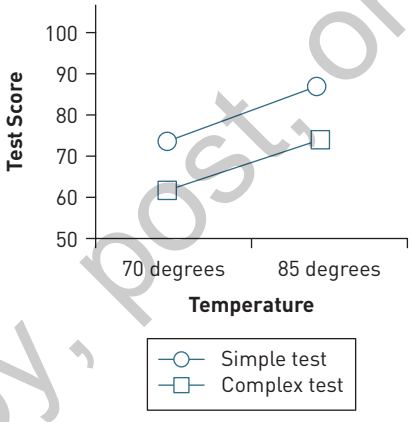
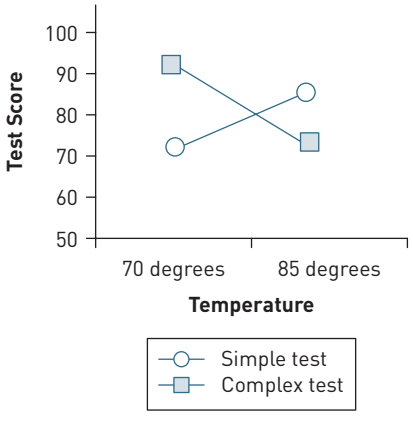


EXHIBIT 8.7 ■ Factorial Plot With Interaction Effects



Advantages of factorial designs include the (1) ability to enhance the "signal" or treatment, (2) potential to examine numerous treatment variations without conducting several sequential experiments, (3) capacity to evaluate interaction effects, and (4) design's abundance of variations to fit different research questions. Among available designs are the 2×2 factorial, a two-factor model where there are more than two levels, the three-factor design, and the incomplete factorial design.[40]

Covariance Analysis and Blocking

There are two types of *noise-reducing experimental designs*: **covariance designs** and **blocking designs**.[41] The analysis of covariance design (ANCOVA or ANACOVA) is a pretest-posttest randomized experimental design. The pretest observation is the same as the posttest but is not required to be identical to the pretest; it can be any variable measured before the intervention. It is also possible for this design to have more than one covariate. The pretest covaries with the outcome measure to accomplish the goal of removing noise, thus the design's name. ANCOVA is the statistical analysis tool used. Regarding "covariates," you might read in a literature review for your project that a "posttest of managerial performance was adjusted for income and educational level." That is, the DV effects were adjusted for these two continuous variables identified as covariates.

The randomized block design is equivalent to the random form of probability sampling: stratification. Like stratification, randomized block designs reduce noise or unexpected variance in the data. This design divides the sample into relatively homogeneous subgroups or "blocks"—like strata. (See sampling plans in Chapter 4.) Each block or homogeneous subgroup becomes a focal point for the experiment. Each block having less variability than the whole sample reduces variability or noise. Estimating the treatment effect across the entire sample is less efficient for performing data analysis; therefore, the block estimates are used. Pooling these more efficient estimates across blocks usually provides an overall estimate better than designs without blocking.[42]

QUASI-EXPERIMENTS

"The prefix *quasi* means 'resembling.' Thus, **quasi-experimental** research is research that resembles experimental research but is not true experimental research."[43] The design specifies a baseline (preintervention) comparison group that tries to achieve a composition as similar to the treatment group as possible. By manipulating the independent variable before the dependent variable, quasi-experiments eliminate any problem with directionality. Nevertheless, *without* random assignment of participants to treatment conditions, confounding variables related to participant selection are not removed. The use of existing or intact groups induces researchers and participants to favor this design for its convenience and relatively less conspicuous and disruptive nature. It is researcher friendly.

Quasi-experiments mimic true experiments except for random assignment, which distinguishes them "because the conclusions that may be drawn from the research depend

In field settings, the researcher has limited leverage over selection but might assign each intact group (a department or team) to either an experimental or control condition. The teams in this photo allow the researcher to have two experimental groups and one control if the experiment calls for it.

upon this distinction. The degree of *risk* in inferring causal relationships is much greater with quasi-experiments."[44] The "quasi-" design is frequently conducted in field settings (industrial, educational, or medical intervention) where random assignment is difficult or impossible. Quasi-experiments resolve issues related to setting, methodology, practical concerns, or ethics that often plague true experiments.

Assignment to conditions occurs using **self-selection** (participants choose the treatment for themselves) or the researcher chooses **intact groups** for assignment. In field settings, the researcher has limited leverage over selection but might assign each *intact group* (a department or team) to either an experimental or control condition. (See the section on "Matching" later in this chapter.) However, the investigator does have influence when controlling the implementation of nonrandom assignment, scheduling observations/measures, equalizing the composition of comparison groups, and affecting some of the features of the treatment.

Example: Nonequivalent Groups Design

In a true experiment, participants in a between-subjects design are randomly assigned to conditions resulting in a similarity of the groups. Indeed, they are considered equivalent. Dissimilarity occurs when participants are not randomly assigned. A **nonequivalent groups design** is a frequently used quasi-experimental between-subjects design in which participants are not equivalent (i.e., not randomly assigned to conditions). The two forms of assignment mentioned previously are self-selecting and intact groups. The latter is used more frequently because it allows the researcher to select different classes in a university, members from similar clubs, or customers from similar stores. The self-selecting form is weaker because participants are recruited and consent to participate based on their interest in being an experimental participant for an undisclosed reason.

In this design, there is a pretest, treatment, and then the dependent variable is measured again using a posttest to see if there is a change from pretest to posttest. The experimental group receives the treatment; the control group receives no treatment while serving as the comparison benchmark. Here, the researcher must consider the literature on the type of treatment and confounding variables known to affect both treatment and instruments/measures, given the limits of nonrandomization.[45] This statement is illustrated in the forthcoming example.

The design has four observations. At each measurement, there may be multiple variables assessed. Two observations (pretests) occur in advance of the treatment, one for

the treatment group and one for the control group. The remaining two occur after the treatment (posttest) for each group, diagramed in Exhibit 8.8.

EXHIBIT 8.8 ■ Nonequivalent Groups Diagram

t_1	t_2	t_3
Pretest	Treatment	Posttest
O_1	X_E	O_2
O_3		O_4
E		
C		

As an example, researchers decide to evaluate teaching SQL to computer science students. SQL is a high-demand programming language that powers many organizations (e.g., businesses, hospitals, banks, universities, etc.) and even Androids and iPhones access SQL databases.[46]

The researchers selected a treatment group composed of one section of a CS400 class and a control group comprising another section of the class. They tried to select groups that were similar; however, in a nonequivalent groups design, student participants self-select into a class of their preference. This self-selection becomes an intact group or class section. The researchers are unaware that some students selected Professor Thompson's class because of his deliberate presentations, whereas high-achieving students selected Professor Collins because she provides practical examples and is more enthusiastic. Neither class section knew that the defining difference in their course curriculum was the presence (or absence) of an SQL programming module.

The experimental group (Collin's class) received five learning modules with the addition of SQL. The control group (Thompson's class) received the same five modules without SQL. Presumably, SQL instruction caused the difference in the two group's programming abilities. Or did it? The differences revealed in Exhibit 8.8 (t_3) may not be due to the experimental treatment at all. Other factors have implications for performance improvement scores, including (1) teaching styles, (2) professor gender, (3) student age, (4) GPA, (5) classroom environment, (6) time of day, (7) student motivation, (8) participation in the student programming club, and (9) breadth of programming experience.

Thus, the researcher should be cognizant of these variables to rule out potential interference. Researchers are not always aware of all of the dimensions on which groups differ, which may be unobservable or unknowable. For causal inference to provide good estimation and be efficient, researchers seek to compare treatment and control groups that are highly similar. As Campbell and Stanley noted regarding this design, "The more similar the experimental and the control groups are in their recruitment, and the more this similarity is confirmed by the scores on the pretest, the more effective this control becomes."[47] If the groups are different, the prediction of the outcome for the control group will be made with information from individuals who are not only distinct from the treatment group but are also different from others in the control group.

If the researchers have thoroughly prepared, the design is said to control for the effects of history, maturation, testing, and instrumentation. Eliminating confounding variables increases the internal validity of this design. Nevertheless, the cautions about intersession history should be taken seriously. For example, there is history (the events other than the treatment

that occur between the pretest and posttest), maturation (participant changes during the time, such as an SQL brown-bag lecture for the department), and regression to the mean (where extreme scores on one occasion tend to move toward the average score the next time).

Other quasi-designs include time series, equivalent materials samples, proxy pretest design, double pretest design, nonequivalent dependent variables design, pattern matching design, and the regression point displacement design.[48]

Matching

Matching is linked to the validity of quasi-experiments and should not be ignored. This warning includes other designs where randomization cannot occur. Methods to compensate for confounding variables depend on what the researcher knows about the experimental and control participants. Matching tries to accomplish the seemingly impossible: to separate out the causal effect from the effects related to preexisting differences between treatment and control groups that were never randomized.[49] **Matching** in many quasi- and field experiments attempts to obtain comparable groups to recreate the feature of randomization; its intent is equivalent or balanced groups. In practice, this involves selecting control group participants using specific criteria (variables) relevant to a study's targeted variables, particularly the DV.

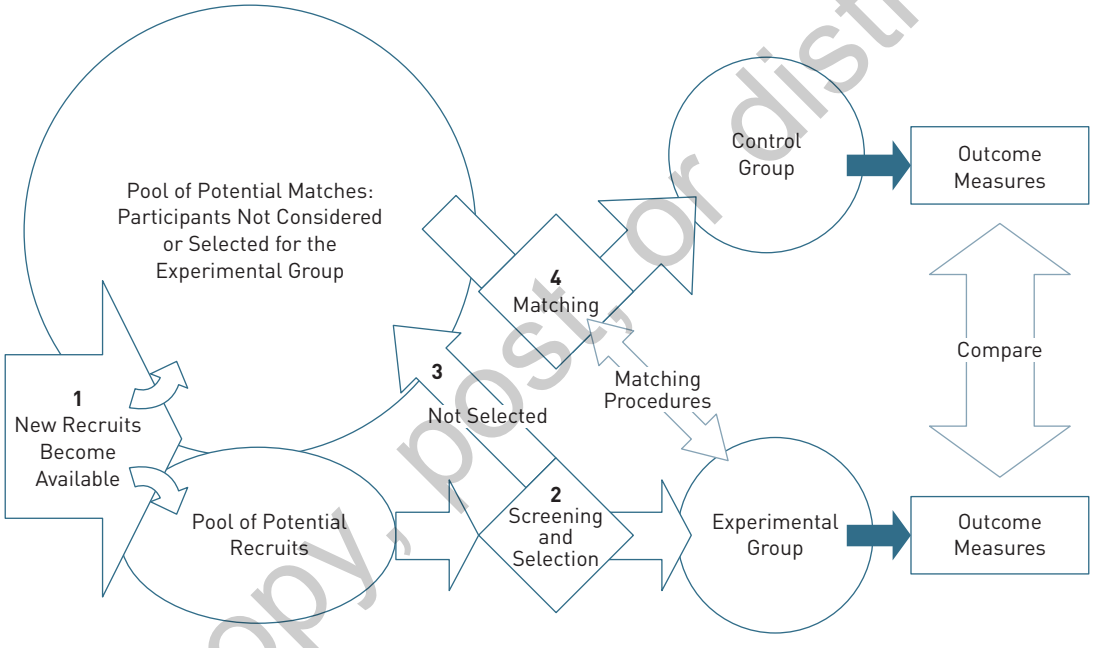
A **quota matrix**[50] is a technique used on a small scale. The matrix represents participant characteristics (variables) affecting the study, which are initially disproportionate because participants recruited for the study are not randomized or evenly distributed in the treatment and control groups. This matrix typically uses categorical variables (e.g., gender, ethnicity, age group) and is reminiscent of a cross-tabulation table. The process ideally results in an even number of participants in each cell so that an allocated number of matched participants are available for selection from the recruitment pool. Additional approaches include stratification, pair matching, and balanced covariates.

Exhibit 8.9 shows another approach. Here, matching is used "to create a control sample, selected from a large donor pool of potential members [so] that the covariate distribution in the matched control group becomes more similar in its covariate distribution to the treated sample."[51] With the quota matrix, discrete or categorical variables are the basis for matching. A **covariate**, on the other hand, is a continuous variable that acts as a control; it is not manipulated but rather is observed and can affect the outcome of the study. A continuous covariate could be education level or test score. In the earlier computer-programming example, a pretest logic score might be a continuous covariate; however, differences among participants on the variables mentioned might overwhelm the instructional SQL treatment.

Follow the numbered steps in the Exhibit 8.9 diagram: (1) → the recruitment produces a pool of potential recruits who are → (2) screened and either selected for the

experimental group or returned to the pool of "potential matches" → (3) because they were unsuitable. The pool of potential matches (minus those not considered/rejected or those already selected for the experimental group) is then → (4) matched using criteria from the experimental group's characteristics and become control group participants. The treatment and measurement of outcomes follow at the "compare" arrow. See the expanded process detail at this reference.[52]

EXHIBIT 8.9 ■ Matching From a Donor Pool of Potential Control Participants



Source: Adapted from Loman, Tony, "Matching Procedures in Field Experiments," *Institute of Applied Research*, 2003, http://capacitybuilding.net/Matching%20Procedures%20in%20Field%20Experiments.pdf.

Other Approaches to Balance Groups

There are several other techniques for balancing treatment and control groups while reducing the effect of and compensating for confounding variables. More advanced methods include regression modeling (of the relationship between the covariates and the outcome measure),[53] nonparametric regression (which has less strict assumptions), distance matching, difference-in-differences, the regression-discontinuity design, and propensity score analysis (modeling the relationship between covariates and treatment assignment). Advanced students will find sources on propensity score analysis at this reference.[54] The **regression-discontinuity design** (RDD) is underused but has

internal validity characteristics that produce inferences comparable to randomized experiments/RCTs and is stronger than nonequivalent groups designs.[55] The RDD is like the pretest-posttest comparison group design but assigns participants to a treatment, or what the designers call a *program*, and to the comparison group using a criterion. The criterion is a cut-off score from the preprogram ("pretest") measure. While a pretest is usually the same instrument administered before and after a treatment, the term "preprogram" thus implies more broadly that before and after measures may be the same or different." The preprogram measure is a continuous variable from which a cut-off or threshold is established, allowing comparisons of observations *close* to either side of the line and thereby estimating an average program effect. Because of the closeness of the scores adjacent to the cut-off, the program group and the control are very similar. As a result, RDD designs are superior to *ex post facto* designs in many ways.[56]

Chapter Summary

- This chapter covered true experiments (sometimes called randomized experiments or randomized control trials), other randomized designs, and quasi-experiments. Pre-experiments were mentioned for comparison purposes only.
- Five questions that companies should answer before beginning a business experiment are as follows: (1) Does the experiment have a clear purpose? (2) Have stakeholders committed to abide by the results? (3) Is the experiment doable? (4) How can we ensure reliable results? (5) Have we gotten the most value out of the experiment?
- The necessary requirements for experiments to make claims of causality are (a) manipulation, (b) control, and (c) random assignment.
- Four essential criteria for establishing causality in the experimental context are as follows: (1) the cause (IV) and effect (DV) are related; (2) the cause precedes the effect (i.e., a time order must be observed); (3) the cause and effect occur together consistently (i.e., cause and effect should *covary*); and (4) alternative explanations can be ruled out: the relationship between the IV and DV must *not* be due to a confounding extraneous "third" variable.
- Threats to validity reduce drawing sound inferences. This chapter reviewed four general categories of validity, provided an internal validity exhibit to help you evaluate your design, and discussed how external validity affects the generalizability of findings.
- True experiments with examples, the need for control groups, and when not to use true experiments were reviewed. Two other categories, other randomized designs (e.g., factorial and covariance analysis/blocking) and quasi-experiments, were also discussed.
- Different matching procedures create equivalent or balanced groups for quasi-experiments. Some of the strategies that attempt to approximate the feature of random assignment were discussed.

Key Terms

block randomization
blocking design
comparison group
confounding variable
construct validity
control
control group
covariance design
covariate
covariate adaptive randomization
criteria for causality
factor
factor level
factorial design
intact group
internal validity
manipulation
matching
nonequivalent groups design
pre-experiment
quasi-experimental
quota matrix
random assignment
random sampling
randomized-based design
regression-discontinuity design
self-selection (to groups)
signal enhancer or noise reducer
simple randomization
statistical conclusion validity
stratified randomization
subject or attribute variable
true experiment

Discussion Questions

1. What five questions should companies answer before committing to a business experiment or be sufficiently knowledgeable about to consider it as a rigorous test for a new product or program launch?
2. Compare and contrast the three major categories of experimental design. Why do some have stronger claims of detecting causality?
3. Discuss the following as essential characteristics of experiments:
   a. Active manipulation of an independent variable. How does the researcher accomplish this?
   b. Researcher control of variables. Discuss why it is essential to control factors that might cause or influence a particular condition *except* those whose possible effects are the focus of the investigation.
   c. Random assignment of participants to groups (or different conditions). How does random assignment assure that (i) each participant has an equal chance of being assigned to a group and (ii) that the assignment of one participant is independent of the assignment of others?
4. What is the difference between random assignment and random sampling? Explain how each is implemented.
5. List four essential criteria for determining causality and explain what must occur to provide assurance that the criterion is met.
6. Differentiate between statistical conclusion validity, internal validity, construct validity, and external validity. Provide examples of each.
7. Describe as many threats to internal validity as you recall and then do the following:
   a. Identify their source.
   b. Describe what remedies are available to the researcher to deal with them.

(Continued)

(Continued)

8. External validity includes population, ecology, temporal, treatment variation, and outcome.
   a. What are the implications of these types of external validity for applying the results of an experiment to a larger population?
   b. Describe the threats to external validity and how they can be managed.
9. Provide an example of a true experimental design.
   a. What is a control group and why is it used?
   b. How do attribute variables defeat random assignment of participants to groups?
10. Other randomized-based designs rely on random assignment to produce group equivalence, to balance treatment and control groups, or to compensate for confounding variables. Describe four types of randomization not including simple randomization.
11. How do quasi-experiments differ from factorial and blocking experiments?
    a. What separates quasi- from true experiments?
    b. Why are quasi-experiments favored for field settings?
    c. Why is matching used in many quasi-experiments?
    d. What are the difficulties for researchers in allowing participants to self-select or to use intact groups?
    e. What measurement level of a variable (NOIR) is used for a quota matrix and how do quota matrices help to match participants in an experiment?
    f. What does "covariate" matching mean and why is it superior to quota matching?
12. Suggest an experimental design for each of the following situations:
    a. A test of salesperson compensation plans where the dependent variable is sales volumes ($) per month. The levels of the IV are as follows:
       (i) Straight salary
       (ii) Salary plus bonus
       (iii) Base plus commission
       (iv) Straight commission
       (v) Variable commission
       (vi) Draw against commission
    b. During a national influenza outbreak, certain people are at higher risk of serious flu complications like respiratory issues requiring hospitalization. A large pharmaceutical company is concerned that their primary antibiotic to treat lung infections is generalized and not targeted to treat co-infections of flu virus and bacteria. They are aware that certain patients are at higher risk of developing severe pneumonia. The company has identified the following categories for testing:
       (i) Children younger than 2 years
       (ii) Adults aged 65 years and older
       (iii) Pregnant women and women up to 2 weeks postpartum
       (iv) Nursing home residents
       (v) People with chronic lung disease

       What other medical conditions/group profiles would you include in the experiment and how would you structure the experiment to discover the effectiveness of the company's drug for specific groups and by level of administered dosage? (Hint: compare your decision to factorial or blocking experiments.)