CHAPTER 1

A PRIMER OF THE SCIENTIFIC METHOD AND RELEVANT COMPONENTS

The primary objective of this book is to help researchers understand and select appropriate designs for their investigations within the field, lab, or virtual environment. Lacking a proper conceptualization of a research design makes it difficult to apply an appropriate design based on the research question(s) or stated hypotheses. Implementing a flawed or inappropriate design will unequivocally lead to spurious, meaningless, or invalid results. Again, the concept of validity cannot be emphasized enough when conducting research. Validity maintains many facets (e.g., statistical validity or validity pertaining to psychometric properties of instrumentation), operates on a continuum, and deserves equal attention at each level of the research process. Aspects of validity are discussed later in this chapter. Nonetheless, the research question, hypothesis, objective, or aim is the primary step for the selection of a research design.

The purpose of a research design is to provide a conceptual framework that will allow the researcher to answer specific research questions while using sound principles of scientific inquiry. The concept behind research designs is intuitively straightforward, but applying these designs in real-life situations can be complex. More specifically, researchers face the challenge of (a) manipulating (or exploring) the social systems of interest, (b) using measurement tools (or data collection techniques) that maintain adequate levels of validity and reliability, and (c) controlling the interrelationship between multiple variables or indicating emerging themes that can lead to error in the form of confounding effects in the results. Therefore, utilizing and following the tenets of a sound research design is one of the most fundamental aspects of the scientific method. Put simply, the research design is the structure of investigation, conceived so as to obtain the "answer" to research questions or hypotheses.

♦ THE SCIENTIFIC METHOD

All researchers who attempt to formulate conclusions from a particular path of inquiry use aspects of the scientific method. The presentation of the scientific method and how it is interpreted can vary from field to field and method (qualitative) to method (quantitative), but the general premise is not altered. Although there are many ways or avenues to "knowing," such as sources from authorities or basic common sense, the sound application of the scientific method allows researchers to reveal valid findings based on a series of systematic steps. Within the social sciences, the general steps include the following: (a) state the problem, (b) formulate the hypothesis, (c) design the experiment, (d) make observations, (e) interpret data, (f) draw conclusions, and (g) accept or reject the hypothesis. All research in quantitative methods, from experimental to nonexperimental, should employ the steps of the scientific method in an attempt to produce reliable and valid results.

The scientific method can be likened to an association of techniques rather than an exact formula; therefore, we expand the steps as a means to be more specific and relevant for research in education and the social sciences. As seen in Figure 1.1, these steps include the following: (a) identify a research problem, (b) establish the theoretical framework, (c) indicate the purpose and research questions (or hypotheses), (d) develop the methodology, (e) collect the data, (f) analyze and interpret the data, and (g) report the results. This book targets the critical component of the scientific method, referred to in Figure 1.1 as *Design the Study*, which is the point in the process when the appropriate research design is selected. We do not focus on prior aspects of the scientific method or any steps that come after the Design the Study step, including procedures for conducting literature reviews, developing research questions, or discussions on the nature of knowledge, epistemology, ontology, and worldviews. Specifically, this book focuses on the conceptualization, selection, and application of common research designs in the field of education and the social and behavioral sciences.

Again, although the general premise is the same, the scientific method is known to slightly vary from each field of inquiry (and type of method). The technique presented here may not exactly follow the logic required for research using qualitative methods; however, the conceptualization of research designs remains the same. We refer the reader to Jaccard and Jacoby (2010) for a review on the various scientific approaches associated with qualitative methods, such as emergent- and discovery-oriented frameworks.

Figure 1.1 The Scientific Method

The Scientific Method
Identify a Need
Establish a Theoretical Foundation
Formulate the Research Question
Design the Study
Collect the Data
Analyze the Data
Report the Results

VALIDITY AND RESEARCH DESIGNS ♦

The overarching goal of research is to reach valid outcomes based upon

... the extent to which the outcome accurately answers the stated research questions of the study. Validity is a complex construct and takes on many different forms, operates on a continuum, and theoretically can be considered multidimensional. In other words, the outcome of most studies cannot typically be dichotomized as valid or not valid. Validity also has a place in psychometrics (i.e., the theories and techniques associated with educational and psychological measurements), and it is generally known as test validity.

The validity of a measurement tool simply means that it measures what it is developed to measure. The focus within this book is the validity related to research designs, *not* test validity (for more information related to test validity, reliability, and measurement, see DeVellis [2011] and Viswanathan [2005]). Although securing validity is critical at the design stage, it should be a consideration throughout the general steps of the scientific method. The importance of securing "acceptable" levels of validity for research in quantitative methods cannot be overstated. However, aspects of validity have also been addressed for qualitative methods. Validity and the qualitative method include focusing in on the *trustworthiness* of the data, such as Lincoln and Guba's (2013) evaluation criteria, as well as the rigor and quality of the data collection procedures (see also Golafshani, 2003; Loh, 2013; Williams & Morrow, 2009). Additionally, the concept of external validity can have a place in qualitative methods as well. We refer the reader to Chenail (2010) for a review on nonprobabilistic approaches to aspects of generalizability for qualitative methods.

In the following sections, we summarize four types of validity related to research designs for quantitative methods: internal, external, construct, and statistical conclusion validity. Originally, the concepts of internal, external, construct, and statistical conclusion validity were all conceptualized for the application and development of experimental and quasi-experimental research (Campbell, 1957; Cook & Campbell, 1979). Since that time, many researchers, books, and Internet references have attempted to classify and order these types of validity very differently in accordance with nonexperimental research, as well as within different disciplines (e.g., epidemiology).

With minor additions, we organize and present the types of validity primarily based on Cook and Campbell's (1979) original work, along with Shadish, Cook, and Campbell's (2002) composition. Any condition that compromises the validity related to a research design is known as a *threat* (i.e., confounding variables). All types of validity are applicable to experimental and quasi-experimental research; however, the conceptualization of internal validity (by definition) does *not* apply to nonexperimental research, including survey and observational (correlational) approaches. Another form of validity—statistical conclusion validity—applies to all research within quantitative methods and refers to the role of statistical analyses and its relation to research design.

Independent and Dependent Variables

In simple terms, the independent variable (IV) is the variable that is manipulated (i.e., controlled) by the researcher as a means to test its impact on the dependent variable, otherwise known as the *treatment effect*. In the classical experimental study, the IV is the treatment, program, or intervention. For example, in a psychology-based study, the IV can be a cognitive-behavioral intervention; the intervention is manipulated by the researcher, who controls the frequency and intensity of the therapy on the subject. In a pharmaceutical study, the IV would typically be a treatment pill, and in agriculture the treatment often is fertilizer. In regard to experimental research, the IVs are always manipulated (controlled) based on the appropriate theoretical tenets that posit the association between the IV and the dependent variable.

Statistical software packages (e.g., SPSS) refer to the IV differently. For instance, the IV for the analysis of variance (ANOVA) in SPSS is the "breakdown" variable and is called a *factor*. The IV is represented as levels in the analysis (i.e., the treatment group is Level 1, and the control group is Level 2). For nonexperimental research that uses regression analysis, the IV is referred to as the *predictor variable*. In research that applies control in the form of statistical procedures to variables that were not or cannot be manipulated, the IVs are sometimes referred to as *quasi-* or *alternate independent variables*. These variables are typically demographic variables, such as gender, ethnicity, or socioeconomic status. As a reminder, in nonexperimental research the IV (or predictor) is not manipulated whether it is a categorical variable such as hair color or a continuous variable such as intelligence. The only form of control that is exhibited on these types of variables is that of statistical procedures. Manipulation and elimination do not apply (see types of control later in the chapter).

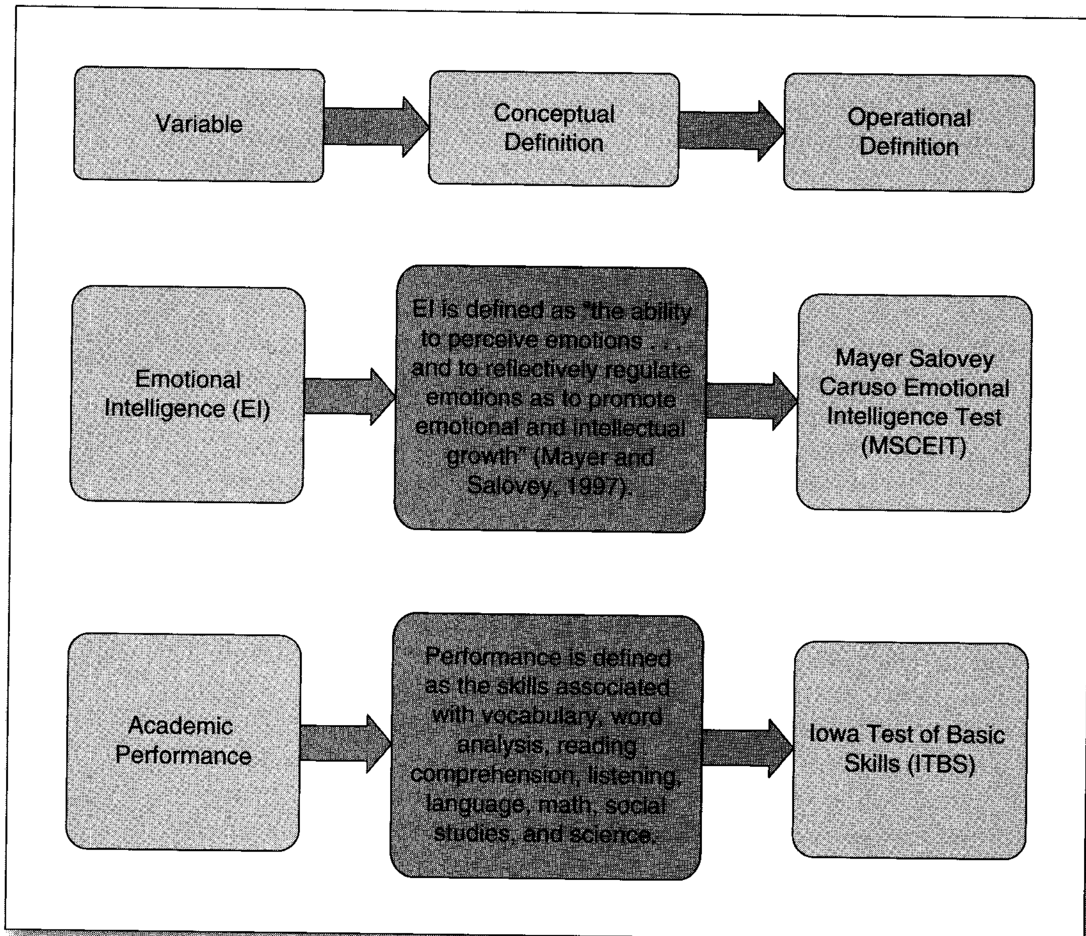
The dependent variable (DV) is simply the outcome variable, and its variability is a function of IV and its impact on it (i.e., treatment effect). For example, what is the impact of the cognitive-behavioral intervention on psychological well-being? In this research question, the DV is psychological well-being. In regard to nonexperimental research, the IVs are not manipuvariables. During the development of research questions, it is critical to first define the DV conceptually, then define it operationally.

A **conceptual definition** is a critical element to the research process and involves scientifically defining the construct so it can be systematically measured. The conceptual definition is considered to be the (scientific) textbook definition. The construct must then be operationally defined to model the conceptual definition.

An **operational definition** is the actual method, tool, or technique that indicates how the construct will be measured (see Figure 1.2).

Consider the following example research question: What is the relationship between *Emotional Intelligence* and conventional *Academic Performance*?

Figure 1.2 Conceptual and Operational Definitions



Internal Validity

Internal validity is the extent to which the outcome was based on the independent variable (i.e., the treatment), as opposed to extraneous or unaccounted-for variables. Specifically, internal validity has to do with causal inferences—hence, the reason why it does not apply to nonexperimental research. The goal of nonexperimental research is to describe phenomena or to explain or predict the relationship between variables, not to infer causation (although there are circumstances when cause and effect can be inferred from nonexperimental research, and this is discussed later in this book). The identification of any explanation that could be responsible for an outcome (effect) outside of the independent variable (cause) is considered to be a threat. The most common threats to internal validity seen in education and the social and behavioral sciences are detailed in Table 1.1. It should be noted that many texts do not indentify *sequencing effects* in the common lists of threats; however, it is placed here, as it is a primary threat in repeated-measures approaches.

Table 1.1 Threats to Internal Validity

Threat	*Explanation*
History	Any event that occurs during the time of the treatment and the posttest that could affect the outcome (e.g., natural life events such as a death in the family, change in job, or moving)
Maturation	The natural process of changing, growing, and learning over time
Testing	The effects of practice familiarity in taking the same test more than once (e.g., the participant who takes the same math achievement test twice in the pretest and posttest measures may improve performance simply because of the familiarity with the test)
Instrumentation	The change in a measuring instrument over time (i.e., some instruments undergo revisions)
Statistical regression	The tendency for any extreme score to regress toward the average (i.e., regression toward the mean is a statistical phenomenon that any extreme scores, high or low, eventually regress or revert to the average)
Selection bias	Also known as *selection effect*; results when researchers do not use a systematic assignment technique (e.g., random assignment) to assign participants to conditions and is the largest threat to internal validity in quasi-experimental research

(Continued)

Table 1.1 (Continued)

Threat	*Explanation*
Attrition	The loss of participants during the term of the experiment (also known as *drop-out* or *subject mortality*)
Combination of selection and other treatments	For designs that include more than one group—any one of the threats to internal validity can affect one of the groups in the study as opposed to the other (e.g., the participants in one condition may have been exposed to a stressful event not related to the experiment, but this event does not affect the other condition)
Diffusion	The inadvertent application of the treatment to the control group (e.g., in educational settings, teachers may use aspects of the math intervention in the control group that are supposed to be delivered only to the control condition)
Special treatment	Special attention to the control group, with the changes attributed only to the attention (i.e., placebo effect)
Sequencing effects –Order effects –Carryover effects	Related to within-subject (repeated-measures) approaches and also known as *multiple-treatment interference, fatigue effects,* and *practice effects*; can be separated into *order effects* (i.e., the order in which participants receive the treatment can affect the results) and *carryover effects* (i.e., performance in one condition affects performance in another condition)

External Validity

External validity is the extent to which the results can be generalized to the relevant populations, settings, treatments, or outcomes. Generally speaking, external validity can be secured if a true probability sampling technique (e.g., random selection) is used, although logistically this is often extremely difficult. Therefore, it is feasible that cause and effect can be established via the application of a sound experiment, but the findings may not generalize to the appropriate population or settings. As seen in Table 1.2, the primary threats to external validity are detailed and primarily slanted toward the examinations of causal relationships. However, issues pertaining to external validity should be considered for nonexperimental research. The most obvious threat to external validity for survey approaches (a form of nonexperimental research), for example, would be *sample characteristics,* sometimes referred to as *sampling bias.*

Table 1.2 Threats to External Validity

Threat	*Explanation*
Sample characteristics	The extent to which the sample (i.e., unit) represents the population from which it is drawn (i.e., for a sample to represent a population, the researcher must employ random selection and the appropriate sampling procedure and power analysis)
Stimulus characteristics and settings	The unique factors involved in providing the treatment or intervention, such as the setting and researchers (i.e., it is difficult to replicate contrived laboratory conditions to real-life scenarios)
Treatment variations	Variations in the same treatment or the combination of multiple or partial treatments that account for different results
Outcome variations	Observing the effect of one type of outcome differs when alternate outcomes are observed
Context-dependent mediation	Mediating variables related to outcomes differ between contexts or settings

Construct Validity

Construct validity refers to the extent a generalization can be made from the operationalization (i.e., the scientific measurement) of the theoretical construct back to the conceptual basis responsible for the change in the outcome. Again, although the list of threats to construct validity seen in Table 1.3 are defined to imply issues regarding cause-effect relations, the premise of construct validity should apply to all types of research. Some authors categorize some of these threats as *social threats* to internal validity, and some authors simply categorize some of the threats listed in Table 1.3 as threats to internal validity. The categorization of these threats can be debated, but the premise of the threats to validity cannot be argued (i.e., a violation of construct validity affects the overall validity of the study in the same way as a violation of internal validity).

Statistical Conclusion Validity

Statistical conclusion validity is the extent to which the statistical covariation (relationship) between the treatment and the outcome is accurate. Specifically, the statistical inferences regarding statistical conclusion validity

Table 1.3 Threats to Construct Validity

Threat	*Explanation*
Attention and contact with participants	Similar to *special treatment;* the level of attention (differentiated attention) from the experimenter varies between the groups (e.g., the researcher spends more time with Group 1 than Group 2, and the differences observed in the outcome can be explained by the increased amount of attention and not due to the intervention)
Single operations and narrow stimulus sampling	The impact the researcher has on the development and implementation of the treatment (i.e., researchers deliver treatments differently based on experiences and expertise; therefore, it is difficult to measure the impact the researcher has on the treatment itself)
Experimenter expectancies	The researchers' expectancies, beliefs, and biases about the results (e.g., if a researcher strongly believes anxiety reduces test performance, then the interaction between the researcher and the participant may influence the outcome because the delivery of instructions and adherence to protocols may change)
Cues of the experimental situation	Sources of influence conveyed to prospective participants (e.g., rumors, information passed along from previous participants)
Novelty effects	The novelty of being in a new or innovative context
Inadequate explication of constructs	The construct under investigation is not appropriately defined conceptually, leading to inadequate measurement (i.e., operationalization)
Construct confounding	Multiple constructs not clearly identified and accounted for operationally
Mono-operation bias	An operationalization (i.e., measurement) does not appropriately represent the construct under investigation, leading to measuring unintended constructs
Mono-method bias	All measurement techniques are the same as a means to measure the construct under investigation
Confounding constructs with levels of constructs	All the levels of a construct are not fully accounted for through the appropriate measurement and reporting tools
Treatment sensitive factorial structure	The interpretation and structure of a measure change as a result of the treatment
Reactivity to assessment	The participants' awareness of being studied may influence the outcome; also known as *acquiescence bias, social desirability,* and the *Hawthorne* or *observer effect*; also an unnatural reaction to any particular form of assessment (i.e., when participants know they are being assessed, the assessment is considered obtrusive and may alter outcome measures other than what they would naturally)

Threat	*Explanation*
Test sensitization	Also known as *pretest sensitization*; the sensitization to the intervention when participants are pretested (e.g., participants are pretested on perceptions of persuasive speeches and are then shown a movie on a persuasive speech; the pretest may influence how they view the speech)
Timing of measurement	The point in time the assessments are administered (i.e., unknown changes may occur, and the different timing of assessments may reveal different results)
Compensatory equalization	When participants in one condition receive more desirable services or compensation compared to that of another condition (thus, constituents may provide enhanced services or goods to the condition not receiving the benefits)
Compensatory rivalry	When participants in the control condition make a concerted effort to make improvements or changes in line with the treatment condition
Resentful demoralization	When participants become resentful or demoralized when they perceive they are receiving a less desirable treatment compared to that of another condition

has to do with the ability with which one can detect the relationship between the treatment and outcome, as well as determine the strength of the relationship between the two. As seen in Table 1.4, the most notable threats to statistical conclusion validity are outlined. Violating a threat to statistical conclusion validity typically will result in the overestimation or underestimation of the relationship between the treatment and outcome in experimental research. A violation can also result in the overestimation or underestimation of the explained or predicted relationships between variables as seen in nonexperimental research.

Table 1.4 Threats to Statistical Conclusion Validity

Threat	*Explanation*
Low statistical power	Power is the extent to which the results of an analysis accurately reveal a statistically significant difference between groups (or cases) when a statistical difference truly exists.
Assumption violation of statistical tests	Violating the assumptions (depending on the extent of the violation) of statistical tests can lead to overestimation or underestimation of practical and statistical significance of an outcome.

(Continued)

Table 1.4 (Continued)

Threat	*Explanation*
Error rate problem	Statistical significance can be artificially inflated when performing multiple pairwise tests; it is also referred to as *family-wise error rate* (i.e., the probability of making a Type I error when performing multiple pairwise analyses).
Restriction of range	A lack of variability between variables weakens the relationship and lowers statistical power.
Extraneous variance in the experimental setting	Variations within the experimental setting (e.g., temperature) may inflate error.
Inaccurate effect size estimation	Some statistical analyses can overestimate or underestimate the size of an effect.
Variability in the procedures	Also referred to as *unreliability of treatment implementation*, the variations in the application of an intervention may affect the outcome (i.e., a nonstandardized approach will create variability in the outcome that is not attributable to the treatment, but rather to the application of the treatment).
Subject heterogeneity	The variability of participant demographics (e.g., age, race, ethnicity, background) may create unaccounted-for variations in the findings.
Unreliability of the measures	Measures maintain certain levels of validity and reliability (pertaining to psychometric principles), and lack of reliability causes inconsistency in measurement.
Multiple comparisons and error rates	The use of multiple dependent variables across conditions and multiple statistical analyses creates greater opportunities for error variance.

The reader is referred to the following books and article for an in-depth review of issues related to validity in research:

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.

Shadish, W. R. (2010). Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods, 15,* 3–17.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

DESIGN LOGIC ♦

The overarching objective of a research design is to provide a framework from which specific research questions or hypotheses can be answered while using the scientific method. The concept of a research design and its structure is, at face value, rather simplistic. However, complexities arise when researchers apply research designs within social science paradigms. These include, but are not limited to, logistical issues, lack of control over certain variables, psychometric issues, and theoretical frameworks that are not well developed. In addition, with regard to statistical conclusion validity, a researcher can apply sound principles of scientific inquiry while applying an appropriate research design but may compromise the findings with inappropriate data collection strategies, faulty or "bad" data, or misdirected statistical analyses. Shadish and colleagues (2002) emphasized the importance of structural design features and that researchers should focus on the theory of design logic as the most important feature in determining valid outcomes (or testing causal propositions). The logic of research designs is ultimately embedded within the scientific method, and applying the principles of sound scientific inquiry within this phase is of the utmost importance and the primary focus of this guide.

Control

Control is an important element to securing the validity of research designs within quantitative methods (i.e., experimental, quasi-experimental, and nonexperimental research). However, within qualitative methods, behavior is generally studied as it occurs naturally with no manipulation or control. Control refers to the concept of holding variables constant or systematically varying the conditions of variables based on theoretical considerations as a means to minimize the influence of unwanted variables (i.e., extraneous variables). Control can be applied actively within quantitative methods through (a) manipulation, (b) elimination, (c) inclusion, (d) group or condition assignment, or (e) statistical procedures.

Manipulation. Manipulation is applied by manipulating (i.e., controlling) the independent variable(s). For example, a researcher can manipulate a behavioral intervention by systematically applying and removing the intervention or by controlling the frequency and duration of the application (see section on independent variables).

Elimination. Elimination is conducted when a researcher holds a variable or converts it to a constant. If, for example, a researcher ensures the temperature in a lab is set exactly to 76° Fahrenheit for both conditions in a biofeedback study, then the variable of temperature is eliminated as a factor because it is held as a constant.

Inclusion. Inclusion refers to the addition of an extraneous variable into the design to test its affect on the outcome (i.e., dependent variable). For example, a researcher can include both males and females into a factorial design to examine the independent effects gender has on the outcome. Inclusion can also refer to the addition of a control or comparison group within the research design.

Group assignment. Group assignment is another major form of control (see more on group and condition assignments later). For the between-subjects approach, a researcher can exercise control through random assignment, using a matching technique, or applying a cutoff score as means to assign participants to conditions. For the repeated-measures approach, control is exhibited when the researcher employs the technique of counterbalancing to variably expose each group or individual to all the levels of the independent variable.

Statistical procedures. Statistical procedures are exhibited on variables, for example, by systematically deleting, combining, or not including cases and/or variables (i.e., removing outliers) within the analysis. This is part of the data-screening process as well. As illustrated in Table 1.5, all of the major forms of control can be applied in the application of designs for experimental and quasi-experimental research. The only form of control that can be applied to nonexperimental research is statistical control.

Table 1.5 Control Techniques for Experimental, Quasi-Experimental, and Nonexperimental Research

Type of Control	*Experimental and Quasi-Experimental Research*	*Nonexperimental Research*
Manipulation	Yes	—
Elimination	Yes	—
Inclusion	Yes	—
Group or condition assignment	Yes	—
Statistical procedures	Yes	Yes

DESIGN NOTATIONS ♦

Design notations are the symbols used to diagrammatically illustrate the process of a research design (see Table 1.6). Within the design, time moves from left to right of the design structure. We used the design notations presented here in each research design covered. The notations presented in this book are based on Campbell and Stanley's (1963) work.

Observation (O). Observation, also known as *measurement*, is symbolized by an "O." The O can refer to a single measure of the dependent variable or multiple measures (O_1, O_2 . . . O_n).

Treatment (X). Treatment, also known as *intervention* or *program* (i.e., the treatment is technically the independent variable and also referred to as a factor), is symbolized with an "X." A control group typically does not receive the treatment and is designated as "-" in its place.

Table 1.6 Design Notations

Design Notation	*Design Element*
O	Observation
X	Treatment
A, B	Factor

Factor (A, B . . . Z). Multiple treatments (factors) used in a design are designated as "X_A" and "X_B" and can go as far up the alphabet as there are factors.

ASSIGNMENT TECHNIQUES ♦

In quantitative methods, each group in a research design has its own line within the structure of the diagram (see Table 1.7). One line equates to one group, two lines equate to two groups, and so on. The assignment of a group is usually the first design notation listed in the line structure.

Random assignment (R). Participants are randomly assigned to each condition to theoretically ensure group equivalency. Logistically, as seen in Figure 1.3, *stratified* random assignment (R_S), sometimes referred

Table 1.7 Group Assignment Design Notations

Design Notation	*Assignment*
R	Random
NR	Nonrandom
C	Cutoff score

the subjects are balanced within predetermined stratum blocks or strata (e.g., age, ethnicity) and then randomly assigned to conditions. See Imgen and Rubin (2015) for more on classical random-assignment approaches, such as Bernoulli trials, completely randomized, stratified, and paired-randomized experiments.

Nonrandom assignment (NR). Participants are assigned to each condition by a matter of convenience or necessity because random assignment is neither an option nor required (nonequivalent groups).

Cutoff score (C). A cutoff score (criterion) is used to assign participants to groups within regression-discontinuity approaches. To create a cutoff criterion, a single pretest continuous distribution is determined and then a division in the data (i.e., cutoff) is made that determines the assignment of participants to conditions.

Matched (M). Matching is a technique used by researchers to match participants on the basis of some extraneous variable that is related to the dependent variable. When this technique is used to assign participants to conditions, some researchers refer to these as match-group designs, but this is not entirely accurate. It is the assignment technique that changes, but the design remains the same.

Matched pairs. For application in any research design indicated in the between-subjects approach, the researcher can (a) match participants in pairs based on certain criteria (e.g., IQ score), then randomly assign each member of the pair to conditions in order to ensure group equivalency (experimental design), and designate this as M_R or (b) match participants based on certain criteria without random assignment to a specific group (quasi-experimental design), then designate this as M_{NR}. For more on matched pairs, see Shadish et al. (2002, p. 118).

Matched grouping. For application in observational approaches, as well as the ex post facto (i.e., after the fact) design, the researcher manually matches participants in groups (M_A) as a means to establish control over the variables of interest. This is conducted because the independent [treatment] variable has already occurred and is not manipulated; therefore, various levels of alternate independent variables (e.g., age, gender) can be statistically manipulated and used as a means to assign individuals to conditions (see more on ex post facto designs later in this guide). This is a form of statistical procedures control often used in epidemiology studies.

Counterbalancing. Counterbalancing is a technique used only in repeated-measures approaches to control for *sequencing effects*. Researchers use counterbalancing to variably expose each group or individual to all the treatments or various treatment levels. The most common form of counterbalancing is conducted at the group level (each group is exposed to the treatment at different sequences). However, counterbalancing can be randomized (sequence is randomly determined for each participant), intrasubject (participants are exposed to more than one sequence, usually in one order, then reversed), complete (every possible sequence is offered), or incomplete (not every sequence is provided because it would require too many conditions, as seen later in the Latin-square design).

The reader is referred to the following article and book for an in-depth review of topics related to group assignment:

Cook, T. D., & Steiner, P. M. (2010). Case matching and the reduction of selection bias in quasi-experiments: The relative importance of pretest measures of outcome, of unreliable measurement, and of mode of data analysis. *Psychological Methods, 15*(1), 56–68.

Rubin, D. B. (2006). *Matched sampling for causal effects.* Cambridge, England: Cambridge University Press.

Figure 1.3 Example of a Stratified Random-Assignment Technique

	Sample of Subjects With GPAs Ranging From 2.0 to 4.0 (N = 52)			
	Subjects With a GPA of 2.0 to 2.5 (n = 14)	*Subjects With a GPA of 2.6 to 3.0 (n = 12)*	*Subjects With a GPA of 3.1 to 3.5 (n = 16)*	*Subjects With a GPA of 3.6 to 4.0 (n = 10)*
	↓	↓	↓	↓
1 Treatment (×)	$n = 7$	$n = 6$	$n = 8$	$n = 5$
2 Control (-)	$n = 7$	$n = 6$	$n = 8$	$n = 5$

Note: This is an example of a two-group design (one treatment and one control group), and the pool of subjects is separated into strata based on grade point average (GPA; i.e., the stratification variable) and then randomly assigned to conditions. Some researchers recommend using this technique when $N < 100$ (Lachin, Matts, & Wei, 1988).

♦ COMPARISON AND CONTROL GROUPS

The group that does not receive the actual treatment, or intervention, is typically designated as the control group. Control groups fall under the *group or condition assignment* aspect of control. Control groups are comparison groups and are primarily used to address threats to internal validity such as history, maturation, selection, and testing. A *comparison group* refers to the group or groups that are not part of the primary focus of the investigation but allow the researcher to draw certain conclusions and strengthen aspects of internal validity. There are several distinctions and variations of the control group that should be clarified.

Control group. The control group, also known as the *no-contact control*, receives no treatment and no interaction.

Attention control group. The attention control group, also known as the *attention-placebo*, receives attention in the form of a pseudo-intervention to control for reactivity to assessment (i.e., the participant's awareness of being studied may influence the outcome).

Nonrandomly assigned control group. The nonrandomly assigned control is used when a no-treatment control group cannot be created through random assignment.

Wait-list control group. The wait-list control group is withheld from the treatment for a certain period of time, then the treatment is provided. The time in which the treatment is provided is based on theoretical tenets and on the pretest and posttest assessment of the original treatment group.

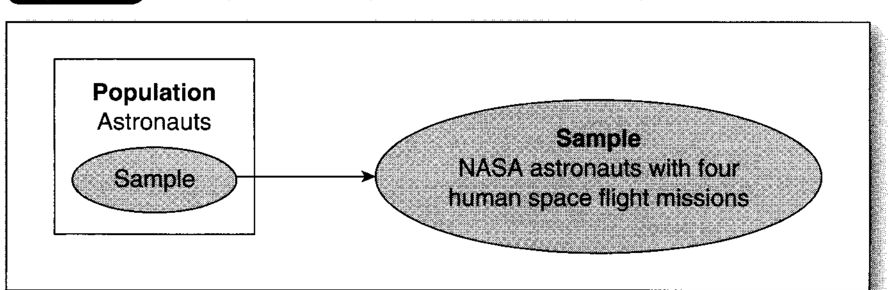
Historical control group. Historical control is a control group that is chosen from a group of participants who were observed at some time in the past or for whom data are available through archival records, sometimes referred to as *cohort* controls (i.e., a homogenous successive group) and useful in quasi-experimental research.

♦ SAMPLING STRATEGIES

A major element to the logic of design extends to sampling strategies. When developing quantitative, qualitative, and mixed methods studies, it is important to identify the individuals (or extant databases) from whom you plan to collect data. To start, the *unit of analysis* must be indicated. The unit of analysis is the level or distinction of an entity that will be the focus of the study. Most commonly, in social science research, the unit of analysis is at the individual or group level, but it can also be at the programmatic level (e.g., institution or state level).

There are instances when researchers identify units nested within an aggregated group (e.g., a portion of students within a classroom) and refer to this as *nested designs* or models. It should be noted that examining nested units is not a unique design, but rather a form of a sampling strategy, and the relevant aspects of statistical conclusion validity should be accounted for (e.g., independence assumptions). After identifying the unit, the next step is to identify the *population* (assuming the individual or group is the unit of analysis), which is the group of individuals who share similar characteristics (e.g., all astronauts). Logistically, it is impossible in most circumstances to collect data from an entire population; therefore, as illustrated in Figure 1.4, a *sample* (or subset) from the population is identified (e.g., astronauts who have completed a minimum of four human space-flight missions and work for NASA).

Figure 1.4 Example of a Sample Extracted From a Population



The goal often, but not always, is to eventually generalize the finding to the entire population. There are two major types of sampling strategies, probability and nonprobability sampling. In experimental, quasi-experimental, and nonexperimental (survey and observational) research, the focus should be on probability sampling (identifying and selecting individuals who are considered representative of the population). Many researchers also suggest that some form of probability sampling for observational (correlational) approaches (predictive designs) must be employed—otherwise the statistical outcomes cannot be genseen in qualitative methods not necessary, some researchers use nonprobability sampling techniques (i.e., the researcher selects participants on a specific criterion and/or based on availability). The following list includes the major types of probability and nonprobability sampling techniques.

Probability Sampling Techniques

Simple random sampling. Every individual within the population has an equal chance of being selected.

Cluster sampling. Also known as *area sampling*, this allows the researcher to divide the population into clusters (based on regions) and then randomly select from the clusters.

Stratified sampling. The researcher divides the population into homogeneous subgroups (e.g., based on age) and then randomly selects participants from each subgroup.

Systematic sampling. Once the size of the sample is identified, the researcher selects every *n*th individual (e.g., every third person on the list of participants is selected) until the desired sample size is fulfilled.

Multistage sampling. The researcher combines any of the probability sampling techniques as a means to randomly select individuals from the population.

Nonprobability Sampling Techniques

Convenience sampling. Sometimes referred to as *haphazard* or *accidental sampling*, the investigator selects individuals because they are available and willing to participate.

Purposive sampling. The researcher selects individuals to participate based on a specific need or purpose (i.e., based on the research objective, design, and target population); this is most commonly used for qualitative methods (see Patton, 2002). The most common form of purposeful sampling is *criterion sampling* (i.e., seeking participants who meet a specific criterion). Variations of purposive sampling include *theory-guided*, *snowball*, *expert*, and *heterogeneity sampling*. *Theoretical sampling* is a type of purposive sampling used in grounded-theory approaches. We refer the reader to Palinkas et al. (2014) for a review of recommendations on how to combine various sampling strategies for the qualitative and mixed methods.

The reader is referred to the following book for an in-depth review of a topic related to sampling strategies for quantitative and qualitative methods:

Levy, P. S., & Lemeshow, S. (2009). *Sampling of populations: Methods and applications* (4th ed.). New York, NY: John Wiley & Sons.

Now that we covered a majority of the relevant aspects to research design, which is the "Design the Study" phase of the scientific method, we now present some steps that will help researchers select the most appropriate design. In the later chapters, we present a multitude of research designs used in quantitative, qualitative, and mixed methods. Therefore, it is important to review and understand the applications of these designs while regularly returning to this chapter to review the critical elements of design control and types of validity, for example. Let's now examine the role of the research question.

RESEARCH QUESTIONS ♦

Simply put, the primary research question sets the foundation and drives the decision of the application of the most appropriate research design. However, there are several terms related to research questions that should be distinguished. First, in general, studies will include an overarching observation deemed worthy of research. The "observation" is a general statement regarding the area of interest and identifies the area of need or concern.

Based on the initial observation, specific variables lead the researchers to the appropriate review of the literature and a theoretical framework is typically established. The purpose statement is then used to clarify the focus of the study, and finally, the primary research question ensues. Research studies can also include hypotheses or research objectives. Many qualitative studies include research aims as opposed to research questions. In quantitative methods (this includes mixed methods), the research question (hypotheses and objectives) determines (a) the population (and sample) to be investigated, (b) the context, (c) the variables to be operationalized, and (d) the research design to be employed.

Types of Inquiry

There are several ways to form a testable research inquiry. For qualitative methods, these can be posed as research questions, aims, or objectives.

while identifying the central phenomenon to be explored. For the application of quantitative methods, researchers can use questions and objectives as well, but also can use hypothesis. Hypotheses are simply predictions the researcher posits as to the direction a relationship will manifest between two or more variables. A hypothesis is purely statistical terminology that is thus tested with statistics. At the heart of every statistical analysis is the null hypothesis. For example, a basic t test is used to examine the mean differences between two groups. The null hypothesis for the t test is that no differences exist between the two groups. The researcher then collects data from the two groups, states an alternate hypothesis to the null, and then analyzes the data with the t test to either reject or accept that null. And in the process, the hypothesis is confirmed or disconfirmed.

Research questions for the quantitative method are still tested in the same manner but are just presented in a different fashion. Creswell's (2014) composition presented three major types of research questions and scripts to be applied to aid in the development of these questions. The three types are the following:

Descriptive. The descriptive question indicates the participants and at least one variable to be investigated. An example could be "What are the anxiety levels of students in the math class?" In this example, the variable to be measured is anxiety levels, and the participants are students in a math class.

Relational. A relationship question includes at least two variables and the participants from which the data should be collected. For example, "What is the relationship between pretest anxiety and test scores for students taking college entrance exams?" The two variables are anxiety and test scores.

Comparison. A comparison question indicates at least two distinct groups and at least one variable that can be measured between the two groups. For example, "How do males compare to females in terms of the their pretest anxiety and test scores on college entrance exams?"

Research questions for the qualitative method are classified as central and subquestions. It is recommended to begin qualitative research questions with open-ended verbs such as *what* or *how* to convey the emerging aspect reflective of the qualitative method.

Central. The central research question is a broad statement of inquiry focused on the exploration of the central or primary phenomenon of focus. For example, a central research question for an ethnographic approach could be "How do Latin-American immigrant children transition into the English-speaking school system?"

Subquestion. The subquestions follow the central question and narrow the focus. The subquestions are a starting point to the development of the qualitative data collection procedures (e.g., interview or focus group questions). Follow-up subquestions, for example, could be "What are the experiences of Latin-American students in the school?" and "How are these experiences reflected at home with their family?"

A flowchart and examples follow that will assist researchers in determining the most appropriate design based on the primary research question of the study. Recall from the Preface the chart that indicated the levels related to determining a design for quantitative and qualitative methods (Method, Research, Approach, and Design). The research question can be broken down, using this chart to determine the most appropriate design.

QUANTITATIVE

Level	Explanation
METHOD$_1$	The *method* provides the theoretical, philosophical, and data analytic stance (e.g., a quantitative method$_1$).
▼	▼
RESEARCH$_2$	*Research* refers to the systematic process of control (e.g., group assignment, selection, and data collection techniques). Research can be experimental, quasi-experimental, or nonexperimental (e.g., a quantitative method$_1$ and experimental research$_2$).
▼	▼
APPROACH$_3$	The *approach* is the first step to creating structure to the design, and it details (a) a theoretical model of how the data will be collected, and (b) if one case, one group, or multiple groups will be associated with the process (e.g., a quantitative method$_1$, experimental research$_2$ with a between-subjects approach$_3$).
▼	▼
DESIGN$_4$	The *design* is the actual structure or framework that indicates (a) the time frame(s) in which data will be collected, (b) when the treatment will be implemented (or not), and (c) the exact number of groups that will be involved (e.g., a quantitative method$_1$, experimental research$_2$ with a between-subjects approach$_3$ and a pre- and posttest control group design.)

QUALITATIVE	
$METHOD_1$	The *method* provides the theoretical, philosophical, and data analytic stance (e.g., a qualitative $method_1$).
▼	▼
RESEARC H_2	*Research* for the qualitative method is nonexperimental (e.g., a qualitative $method_1$ and nonexperimental $research_2$).
▼	▼
$PERSPECTIVE_3$	The *perspective* is the first step to creating structure to the design, and it details the theoretical perspective (or lens) of how the researcher(s) will spproach the study (e.g., a qualitative $method_1$, nonexperimental $research_2$ with an ethnographic $perspective_3$).
▼	▼
$DESIGN_4$	The *design* is the actual structure that indicates (a) if one case, one group, or multiple groups will be a ssociated with the process, and (b) when the data will be analyzed (e.g., a qualitative $method_1$, nonexperimental $research_2$ with an $ethnographic_3$ and a case study $design_4$).

Type of Research Question	
Level	*Question*
$METHOD_1$	Quantitative or qualitative
▼	▼
$RESEARCH_2$	Experimental, quasi-experimental, or nonexperimental
▼	▼
$APPROACH_3$	Quantitative or qualitative methodological variant
▼	▼
$DESIGN_4$	Any design variant found under the quantitative or qualitative method

Example 1.1

Descriptive	
Level	What are the levels of perceived anxiety students experience prior to testing?
$METHOD_1$	Quantitative
▼	▼
$RESEARCH_2$	Nonexperimental

Descriptive	
▼	▼
$APPROACH_3$	Survey
▼	▼
$DESIGN_4$	Cross-sectional

Note: Perceived anxiety is the only variable in this question that requires operationalization. It is likely that a cross-sectional design will suffice, but if time allows for it, a longitudinal design can be employed.

Example 1.2

Relational	
Level	To what extent do levels of perceived anxiety predict performance on standardized testing?
$METHOD_1$	Quantitative
▼	▼
$RESEARCH_2$	Nonexperimental
▼	▼
$APPROACH_3$	Observational
▼	▼
$DESIGN_4$	Predictive

Note: The variables in this question are anxiety and test performance. This is a relational question that qualifies as an observational approach. The design can be explanatory, but if the data points are not collected at the same time (i.e., anxiety collected at Time Point 1 and then test performance at Time Point 2), then a predictive form of analysis can be used to reduce the data for further interpretation and discussion.

Example 1.3

Comparison	
Level	How do the groups differ between the high-anxiety and low-anxiety conditions in terms of test performance?
$METHOD_1$	Quantitative
▼	▼
$RESEARCH_2$	Experimental

Example 1.3 (Continued)

Comparison	
▼	▼
$APPROACH_3$	Between-subjects
▼	▼
$DESIGN_4$	2-factor posttest

Note: The research question includes one outcome variable broken down into two levels (high and low anxiety). This would require two groups to examine the differences. If random assignment to conditions is employed, then the research is experimental and only a 2-factor posttest design can be employed. If enough participants are available, a third group can be included and considered a control group. If time is on the researcher's side, then a pretest can be included as well, but it is not necessary, particularly if random assignment to conditions is employed.

Example 1.4

Comparison	
Level	How do the groups differ when exposed to the high-anxiety and low-anxiety conditions in terms of test performance?
$METHOD_1$	Quantitative
▼	▼
$RESEARCH_2$	Experimental
▼	▼
$APPROACH_3$	Within-subjects
▼	▼
$DESIGN_4$	2-factor crossover

Note: Similar to the previous example, there is one outcome (dependent) variable at two levels. However, if the researcher has access to only a small group of participants, then a within-subjects (repeated-measures) approach can be used. The participants would experience both conditions through the application of the 2-factor crossover design.

Example 1.5

Central Question	
Level	What are the experiences of parents who have children diagnosed with a pervasive developmental disorder (PDD)?
$METHOD_1$	Qualitative
▼	▼
$RESEARCH_2$	Nonexperimental

Central Question	
▼	▼
PERSPECTIVE$_3$	Narrative
▼	▼
DESIGN$_4$	Descriptive

Note: The central phenomenon is the experience of parents who have children with PDDs. In this example, the researcher is interested in using the narrative perspective as a means to simply provide storytelling to understand the phenomenon. The descriptive design further delineates the perspective that the goal is to provide the narrative of the life stories without providing a critique or assuming there are causes for the resulting phenomenon.

Example 1.6

Central Question	
Level	What are the instructional approaches used by instructors to deal with multicultural populations in graduate school?
METHOD$_1$	Qualitative
▼	▼
RESEARCH$_2$	Nonexperimental
▼	▼
PERSPECTIVE$_3$	Ethnographic
▼	▼
DESIGN$_4$	Realist

Note: The phenomenon to be explored is the instructional approaches for multicultural populations. The ethnographic perspective is adequate in that it will guide the researcher to further understand the point of view of participants from varied cultural backgrounds. The instructional approaches can be culled down for reporting as guided through the realist design.

Keep in mind the examples only reflect general guidelines. Often, researchers pose multiple research questions, which are considered spinoffs of the primary questions. Although this doesn't change the research design, it guides the type of analysis required to properly interpret the data. In summary, if the primary question is descriptive, then the research will be nonexperimental, and a survey approach should be employed. If the primary question is comparative, then any approach and design that falls under the predictive or explanatory design should be applied. As a reminder, the application of the appropriate design relative to the primary research can vary depending on the specific research scenario and the field from which the examination is to be applied. The reader is referred to White (2009) for an in-depth review of the development of research questions for social scientists.

Reviewing the Content and Testing Your Knowledge

Discussion Points

1. Explain from a technical viewpoint why it is important to distinguish a method, research, approach, and design. Next, briefly discuss how understanding each term individually in addition to how these terms interconnect is important for your understanding of the application of research designs.
2. Discuss the importance of validity and research design. Next, choose one type of validity (internal, external, construct, or statistical conclusion) and discuss its relevance to experimental, quasi-experimental, and nonexperimental research.

Exercise

1. Define a sampling strategy.
2. Define the two major types of sampling strategies.
3. Identify a hypothetical population.
4. Identify the sample.
5. What type of sampling strategy will be used?
   a. Why did you choose this type of strategy?
6. Based on the strategy, what type of sampling technique will be used to identify the sample?
   a. Why did you choose this type of technique?

PART I

Quantitative Methods for Experimental and Quasi-Experimental Research

Part I includes four popular approaches to the quantitative method (experimental and quasi-experimental only), followed by some of the associated basic designs (accompanied by brief descriptions of published studies that used the design). Visit the companion website at **study.sagepub.com/edmonds2e** to access valuable instructor and student

Figure I.1 Quantitative Method Flowchart

Method	*Quantitative*				
Research	Experimental and Quasi-Experimental				
Approach	Between-Subjects	Within-Subjects	Regression Discontinuity	Single-Case	
Design	Pretest-Posttest	Posttest	Factorial	Solomon *N*-Group	A-B

resources. These resources include PowerPoint slides, discussion questions, class activities, SAGE journal articles, web resources, and online data sets.

Research in quantitative methods essentially refers to the application of the systematic steps of the scientific method, while using quantitative properties (i.e., numerical systems) to research the relationships or effects of specific variables. Measurement is the critical component of the quantitative method. Measurement reveals and illustrates the relationship between quantitatively derived variables. Variables within quantitative methods must be, first, conceptually defined (i.e., the scientific definition), then operationalized (i.e., determine the appropriate measurement tool based on the conceptual definition). Research in quantitative methods is typically referred to as a *deductive process* and iterative in nature. That is, based on the findings, a theory is supported (or not), expanded, or refined and further tested.

Researchers must employ the following steps when determining the appropriate quantitative research design. First, a measurable or testable research question (or hypothesis) must be formulated. The question must maintain the following qualities: (a) precision, (b) viability, and (c) relevance. The question must be *precise* and well formulated. The more precise, the easier it is to appropriately operationalize the variables of interest. The question must be *viable* in that it is logistically feasible or plausible to collect data on the variable(s) of interest. The question must also be *relevant* so that the result of the findings will maintain an appropriate level of practical and scientific meaning. The second step includes choosing the appropriate design based on the primary research question, the variables of interest, and logistical considerations. The researcher must also determine if randomization to conditions is possible or plausible. In addition, decisions must be made about how and where the data will be collected. The design will assist in determining when the data will be collected. The unit of analysis (i.e., individual, group, or program level), population, sample, and sampling procedures should be identified in this step. Third, the variables must be operationalized. And last, the data are collected following the format of the framework provided by the research design of choice.

♦ EXPERIMENTAL RESEARCH

Experimental research (sometimes referred to as *randomized experiments*) is considered to be the most powerful type of research in determining causation among variables. Cook and Campbell (1979) presented three conditions that must be met in order to establish cause and effect:

1. **Covariation** (the change in the cause must be related to the effect)
2. **Temporal precedence** (the cause must precede the effect)
3. **No plausible alternative explanations** (the cause must be the only explanation for the effect)

The essential features of experimental research are the sound application of the elements of control: (a) manipulation, (b) elimination, (c) inclusion, (d) group or condition assignment, or (e) statistical procedures. Random assignment (not to be confused with random selection) of participants to conditions (or random assignment of conditions to participants [counterbalancing] as seen in repeated-measures approaches) is a critical step, which allows for increased control (improved internal validity) and limits the impact of the confounding effects of variables that are not being studied.

The random assignment to each group (condition) theoretically ensures that the groups are "probabilistically" equivalent (controlling for selection bias), and any differences observed in the pretests (if collected) are considered due to chance. Therefore, if all threats to internal, external, construct, and statistical conclusion validity were secured at "adequate" levels (i.e., all plausible alternative explanations are accounted for), the differences observed in the posttest measures can be attributed fully to the experimental treatment (i.e., cause and effect can be established). Conceptually, a causal effect is defined as a comparison of outcomes derived from treatment and control conditions on a common set of units (e.g., school, person).

The strength of experimental research rests in the reduction of threats to internal validity. Many threats are controlled for through the application of random assignment of participants to conditions. *Random selection,* on the other hand, is related to sampling procedures and is a major factor in establishing external validity (i.e., generalizability of results). Randomly selecting a sample from a population would be conducted so that the sample would better represent the population. However, Lee and Rubin (2015) presented a statistical approach that allows researchers to draw data from existing data sets from experimental research and examine subgroups (post hoc subgroup analysis). Nonetheless, random assignment is related to design, and random selection is related to sampling procedures. Shadish, Cook, and Campbell (2002) introduced the term *generalized causal inference.* They posit that if a researcher follows the appropriate tenets of experimental design logic (e.g., includes the appropriate number of subjects, uses random selection and random assignment) and controls for threats of all types of validity (including test validity), then valid causal inferences can be realized once multiple replications of the experiment are conducted and comparable results can be observed over time (replication being the operative word). Though, recently there have been concerns related to the reproducibility of experimental studies published in the field of psychology, for example (see Baker, 2015; Bohannon, 2015).

Reproducibility could be enhanced if the proper tenets of the scientific method are followed and the relevant aspects of validity are addressed (i.e., internal and construct). Researchers tend to gloss over these constructs and rarely report how they ensured the data to be valid, often assuming that a statistical analysis could be used to "fix" or overshadow the inherent problems of the data. Bad data is clearly the issue, which lends to a great computer science saying "Garbage in, garbage out." To be more specific, taking the appropriate measures to ensure design and test validity, the data will be more "clean," which results in fewer reporting errors in the statistical results. Although probability sampling (e.g., random selection) adds another logistical obstacle to experimental research, it should also be an emphasis along with the proper random assignment techniques.

Although this book is more dedicated to the application of research designs in the social and behavioral sciences, it is important to note the distinction between research designs in the health sciences to that of the social sciences. Experimental research in the health or medical sciences shares the same designs, although the terminology slightly differs, and the guidelines for reporting the data can be more stringent (e.g., see Schultz, Altman, & Moher, 2010, and Appendix H for guidelines and checklist). These guidelines are designed to enhance the quality of the application of the design, which in turn leads to enhanced reproducibility. The most common term used to express experimental research in the field of medicine is randomized control trials (RCT). RCT simply infers that subjects are randomly assigned to conditions. The most common of the RCT designs is the *parallel-group approach*, which is another term for the *between-subject approach* and is discussed in more detail in the following sections. RCTs can also be crossover and factorial designs and are designated under the within-subjects approach (repeated measures).

♦ QUASI-EXPERIMENTAL RESEARCH

The nonrandom assignment of participants to each condition allows for convenience when it is logistically not possible to use random assignment. Quasi-experimental research designs are also referred to as *field research* (i.e., research is conducted with an intact group in the field as opposed to the lab), and they are also known as *nonequivalent designs* (i.e., participants are not randomly assigned to each condition; therefore, the groups are assumed nonequivalent). Hence, the major difference between experimental and quasi-experimental research designs is the level of control and assignment to conditions. The actual designs are structurally the same, but the analyses of the data are not. However, some of the basic pretest and posttest designs can be modified (e.g., addition of multiple observations or inclusion of comparison groups) in an attempt to compensate for lack of group equivalency. In the design structure, a dashed line (- - -) between groups indicates the participants were not randomly assigned to conditions. Review Appendix A for more examples of "quasi-experimental" research designs (see also the example of a diagram in Figure 1.2).

Because there is no random assignment in quasi-experimental research, there may be confounding variables influencing the outcome not fully attributed to the treatment (i.e., causal inferences drawn from quasi-experiments must be made with extreme caution). The pretest measure in quasi-experimental research allows the researcher to evaluate the lack of group equivalency and selection bias, thus altering the statistical analysis between experimental and quasi-experimental research for the exact same design (see Cribbie, Arpin-Cribbie, & Gruman, 2010, for a discussion on tests of equivalence for independent group designs with more than two groups).

Figure 1.2 Double Pretest Design for Quasi-Experimental Research

Group	*Assignment*	*Pretest$_1$*	*Pretest$_2$*	*Treatment*	*Posttest*
1	NR	O_1	O_2	X	O_3
2	NR	O_1	O_2	—	O_3

Time ▶

Note: This is an example of a between-subjects approach with a double pretest design. The double pretest allows the researcher to compare the "treatment effects" between O_1 to O_2, and then from O_2 to O_3. A major threat to internal validity with this design is testing, but it controls for selection bias and maturation. The two pretests are not necessary if random assignment is used.

It is not recommended to use posttest-only designs for quasi-experimental research. However, if theoretically or logistically it does not make sense to use a pretest measure, then additional controls should be implemented, such as using historical control groups, proxy pretest variables (see Appendix A), or the matching technique to assign participants to conditions.

The reader is referred to Shadish, Clark, and Steiner (2008) for an in-depth discussion of how to use linear regression and propensity scores to approximate the findings of quasi-experimental research to experimental research. They discuss this in the greater context of the potential weaknesses and strengths of quasi-experimental research in determining causation.