# 2 THE RESEARCH PROCESS



## LEARNING OBJECTIVES: FAST FACTS

2.1 The two major research approaches—nonexperimental and experimental—offer different yet complementary methods to investigate a question, to evaluate a theory, and to test a hypothesis.

2.2 Both approaches can call upon a toolbox of methods and study designs, and all are evaluated on the basis of reliability and validity.

2.3 Science focuses only on peer-reviewed research, which is evaluated using principles of reliability and validity.

2.4 Cross-cultural research reinforces the critical importance of reliability and validity in evaluating evidence and in optimizing the generalizability of psychological knowledge.

## TESTING BEFORE LEARNING

Here is the answer. You come up with the question.

1. One or more independent variables are deliberately manipulated to assess their effects on one or more dependent variables.
   a. What is a survey study?
   b. What is an ethnographic study?
   c. What is an experimental group?
   d. What is an experiment?
2. A defining feature of a true experiment.
   a. What is random selection?
   b. What is external validity?
   c. What is random assignment?
   d. What is a baseline?
3. A research study compares levels of depression for participants, selected on the basis of a diagnosis of posttraumatic stress syndrome, with a control group.
   a. What is a quasi-experiment?
   b. What is a true experiment?
   c. What is a survey study?
   d. What is a treatment group?
4. The best research design to test cause-and-effect relationships.
   a. What is a quasi-experiment?
   b. What is a randomized experiment?
   c. What is external validity?
   d. What is a survey study?
5. An essential criterion for scientific literature.
   a. What is an impact factor?
   b. What is peer review?
   c. What is an experiment?
   d. What is an independent variable?
6. Cross-cultural research examines the generalizability of evidence.
   a. What is internal validity?
   b. What is reliability?
   c. What is external validity?
   d. What is convergence?

## CAN MONEY BUY YOU HAPPINESS?

Can money buy happiness? So begins a report on a set of research studies conducted by Elizabeth W. Dunn, Lara B. Aknin, and Michael I. Norton, published in the prestigious journal *Science* in 2008. You might intuitively respond "yes" to this question, believing, of course, more money would make anyone happier. Your answer would put you in the company of most people who believe that they would be happier if they were richer. But as we learned in Chapter 1, our intuitive beliefs can often deceive us, and if you answered yes, you would be wrong. This is because scientific studies show that income has only a small effect

on happiness within nations (Kahneman, Krueger, Schkade, Schwarz, & Stone, 2006), particularly in countries with adequate resources, such as housing, sanitation, education, and health care (Diener, Nickerson, Lucas, & Sandvik, 2002). In fact, surveys in multiple countries indicate that levels of happiness have remained unchanged over the last several decades, even though real incomes have dramatically increased (Kahneman et al., 2006).

Intrigued by this surprising and counterintuitive finding, Dunn and colleagues sought to examine why income has such a weak effect on happiness. Research studies offer several explanations, all of which likely contribute in different degrees. For example, researchers have argued that happiness is driven by relative income rather than real income (e.g., R. Frank, 2004). In other words, your happiness is linked to how much you earn in relation to others, such as friends and peers, rather than your absolute income. This we can call the relative income hypothesis of happiness. This hypothesis would also explain why happiness has not increased as nations have grown richer. As the overall wealth of a nation increases, relative income, on average, remains unchanged.

Dunn and colleagues offered another interesting take on this curious question of happiness and money. They proposed that how people use their increased wealth—what activities they pursue and what products they consume—is important for understanding why money by itself does not buy happiness. They cite a 2004 paper appearing in *Daedalus*, "How Not to Buy Happiness," in which the author, Cornell College economics professor Robert H. Frank, argues that people in affluent societies use their income gains on pursuits that offer little in the way of enduring happiness, such as buying bigger houses and purchasing more expensive cars.

Dunn and colleagues thus developed a very exciting set of studies to examine how spending choices might influence happiness. They proposed a novel hypothesis that *prosocial spending*—that is, spending money on others—would be an important factor contributing to happiness. As we will learn in this chapter, their research studies used diverse methods and techniques to design investigations to test their prosocial spending hypothesis of happiness. We will learn how to define, describe, and design these and other related studies in this chapter. Crucially, we will learn the value of studies that yield similar results across diverse methodologies. Such convergence provides the strongest evidence in support of a research hypothesis. For Dunn and colleagues, their very exciting converging evidence supported their hypothesis that prosocial spending promotes happiness. Giving makes you happy! Now let us learn about the diversity of methods that will allow us to conduct similarly exciting research.

## RESEARCH APPROACH

The mindset of a psychological researcher begins with a question about an idea, concept, or theory that can be addressed via the scientific method. For example, a cognitive psychologist might study learning and memory, a clinical psychologist might study schizophrenia, a developmental psychologist might study risk taking in adolescence, and a social psychologist might study racial bias. And, of course, the Dunn et al. study, covered above, asked the age-old question as to whether money can buy happiness.

All these studies will be grounded in the scientific method, meaning that researchers will collect empirical evidence that can be used to answer a particular research question.

These questions can range from whether a specific treatment works for persons diagnosed with mental illness, to why adolescence is a period of heightened risk taking, to how implicit attitudes might influence racial bias, to how happy you are. Scientific studies ideally provide empirical evidence to answer these questions. This empirical evidence comes in the form of quantitative results. These results are then interpreted in light of the research question being asked by the study.

How is such research done? How does a researcher use the scientific method in designing a study? These are questions that we focus on in this chapter. Here, we present the idea of a *research toolbox*. This term is used metaphorically to capture some of the variety of methods and tools researchers use to investigate scientific questions. The methods and tools you choose will depend greatly on the questions asked and the answers you hope to discover. Some research questions may be best studied under highly controlled laboratory conditions. For this type of study, you might directly manipulate your independent variable and carefully measure its effect on a dependent variable. For other research questions, laboratory-based studies may not capture the essence of the topic under investigation. Oftentimes, however, researchers will design a set of studies using different kinds of methods to examine the same question, as was the case in the Dunn et al. research.

There are two very broad research approaches: **experimental** and **nonexperimental**. Nonexperimental studies are common in psychology, and cover a range of investigations that focus on one or more specific variables, such as intelligence, anxiety, learning, and memory. In some instances, the question may be simply to examine the range or distribution of scores on a measure of intelligence, anxiety, or happiness for a particular group of individuals. As an example of a nonexperimental approach, Dunn and colleagues conducted a nationally representative survey study asking 632 Americans about their happiness, charitable donations, and annual incomes.

Other nonexperimental approaches may focus on the *relationship* between two variables, say anxiety and intelligence or happiness and charitable donations. The research question might be whether anxiety detracts from intelligence or whether spending on others promotes happiness. As we will learn, the chief advantage of a nonexperimental approach is that it allows for the quantitative examination and comparison of variables that cannot be directly manipulated. That is, a nonexperimental approach can tell you critical information about the distribution of a variable in a sample or the extent to which two or more variables are related. However, a nonexperimental approach cannot establish cause and effect. This is a crucial point: A nonexperimental approach may show an association of higher anxiety with lower intelligence scores, but you cannot conclude that anxiety caused intelligence to be lower.

On the other hand, with an experimental approach, a researcher deliberately chooses one or more independent variables to be manipulated and then measures the effects of this manipulation on a dependent variable. For example, Dunn et al. incorporated an experimental approach to test their prosocial spending hypothesis of happiness. As described below, they examined whether giving participants either $5 or $20 and instructing them to spend this money on either themselves or others would affect the participants' reported levels of happiness. In other words, Dunn and colleagues directly manipulated two important variables: the amount of money given to participants, and instructions that directed participants to spend this money on either themselves or someone else. They then, as their dependent variable, measured participants' level of happiness.

The independent variable is the presumed cause that produces the effect captured by the dependent variable. In the Dunn et al. study, there were two independent variables that were hypothesized to have a causal effect on happiness. In everyday life, we regularly conduct our own personal experiments, as, for instance, when we assess what happens to our weight if we consume fewer calories, to our blood pressure if we exercise more, or to our grades if we refrain from checking our smartphones during class time! However, as we will learn, there are key research design features that define an experimental approach. As we will also learn, the crucial feature that makes a research approach *experimental* is the use of *random assignment* to create treatment and control groups.

The principal objective of all different kinds of psychological research is to provide a scientific understanding of the topic of investigation. To do so, a research study must outline careful and precise definitions of key concepts and how they will be measured. Generally, researchers describe and define terms both conceptually and operationally. A **conceptual definition** provides the meaning, often rather broad in scope, of an abstract term, such as *intelligence*, *anxiety*, or *emotion*. Very similar to what you would find in a dictionary, a conceptual definition demarcates a semantic or linguistic meaning of a psychological term—that is, its usage in words, texts, and language. For example, intelligence as a concept may be defined as the general ability that enables an individual to comprehend the world and to deal effectively with its challenges (Wechsler, 1997). Or, as in the case of Dunn and colleagues, they defined happiness, conceptually, as a sense of subjective well-being that comes with feeling satisfied with one's life. They defined prosocial spending, conceptually, in terms of the amount of money that a person uses to help others, such as by donating to charity.

Operational definitions follow from conceptual definitions. An **operational definition** indicates how a concept is coded, measured, or quantified. It may be as simple as an operational definition of gender in which *female* is coded as 1 and *male* as 2 (or vice versa). No single operational definition can capture fully the concept it is intended to measure. An operational definition is among several possible objective and measurable indicators of a concept. For Dunn and colleagues, they asked their research participants to estimate how much they spent in a typical month on (a) bills and expenses, (b) gifts for themselves, (c) gifts for others, and (d) donations to charity. The researchers then computed a *personal spending* index, which equaled the total sum of money spent in a typical month on bills and expenses plus gifts for themselves (personal spending = bills and expenses + gifts for themselves), and a *prosocial spending* index, which equaled the total sum of money spent in a typical month on gifts for others plus donations to charity (prosocial spending = gifts for others + donations to charity). Each of these indexes served as an operational definition of personal spending and prosocial spending, respectively.

## Research Strategies

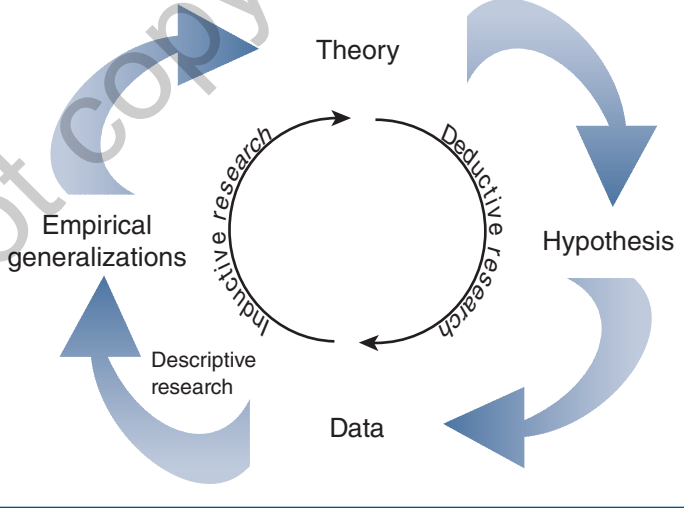
In conducting research, we are attempting to connect theory with empirical data—evidence we obtained through scientific studies. As we have learned, researchers may make this connection by starting with a psychological theory and testing some of its implications with data. This is the process of **deductive research**; it is most often the strategy used in experimental studies. We deduce a hypothesis from a theory and then collect data with which to test the hypothesis. Alternatively, researchers may develop a connection between psychological theory and data by first systematically collecting data, identifying patterns in the data, and then developing a theory that explains the patterns. This **inductive research**

approach is most often used in nonexperimental studies, such as those that use naturalistic observation methods. A particular research study can draw on both inductive and deductive strategies.

So the two most important elements of all scientific research strategies are (a) data and (b) theory. Data are the empirical observations that allow for evaluating a theory. A theory is a set of propositions that interprets and explains the empirical observations of a study. A theory performs three major functions: organization, explanation, and prediction. A good theory is one that is parsimonious and precise as well as powerful in its breadth and depth of explanation. It can explain a variety of occurrences with the fewest theoretical assertions.

In reality, the process of conducting research designed to test explanations for psychological phenomena involves a dynamic interplay of moving from theory to data and then back to theory. This process can be characterized as a **research circle**. As Exhibit 2.1 shows, deductive and inductive research processes can be closely intertwined. With deductive research, theory gives birth to hypotheses, which are then tested by data. With inductive research, data give birth to an **empirical generalization**—a statement that describes patterns found in the data, from which a theory is formulated. (A **generalization** is a broad statement that cannot be directly tested, but rather needs to be translated into one or more hypotheses.) The goals of inductive and deductive research approaches are identical: to develop and formulate theories, a set of general propositions that serve to organize and interpret data or to generate predictions for events and actions that are measured with data.

**EXHIBIT 2.1 ■ The Research Circle**




*Source:* Schutt, R. K. (2012). *Investigating the social world: The process and practice of research* (7th ed., p. 41). Thousand Oaks, CA: Pine Forge Press.

## RESEARCH TOOLBOX

In this book, we will learn about a variety of research tools that psychologists use to investigate scientific questions (see Exhibit 2.2). Each of these tools has its own advantages and disadvantages, and each will be judged by the extent to which it can reliably and validly address the research question posed by a study. Below we introduce some of the different types of methods commonly used in psychological research, as well as some of the different types of study designs. Oftentimes, a research study will incorporate different techniques, combining, for example, survey measures with interviews and observational rating scales. Or, as we will see, the Dunn et al. research program included a nonexperimental study with survey techniques along with two experimental investigations.

### Randomized "True" Experiment

An **experiment** is a study in which one or more independent variables are deliberately manipulated to assess their effects on one or more dependent variables. Independent variables will differ in the degree to which they can be controlled or manipulated. We will learn that in a "true" experiment, the researcher controls *who*, *what*, and *how* the study is conducted. That is, a researcher designs an experiment in which a particular

**EXHIBIT 2.2 ■ Research Toolbox**

| | |
|---|---|
| **True Experiments** | Random assignment of participants to groups and manipulation of one or more independent variables. (Chapters 8 and 9) |
| **Quasi-Experiments** | Experiments in which random assignment of participants to groups is not possible. (Chapter 11) |
| **Correlational Research** | Studies that focus on the distribution of variables, the quantitative association of variables; causation cannot be established. (Chapter 7) |
| **Sampling and Survey Design** | Research in which information is obtained from a sample of individuals through their responses to specific questions. (Chapters 6 and 13) |
| **Performance-Based Measures** | Studies of data collected from standardized tests. (Chapter 5) |
| **Literature Review** | Integrative review of empirical studies that seeks to summarize past research by drawing overall conclusions from many separate investigations that address related or identical hypotheses. (Chapter 3) |
| **Qualitative Research** | Qualitative methods, such as participant observation, intensive interviewing, and focus groups, used to study and understand phenomena in terms of the meanings people attach to them. (Chapter 14) |

aspect of the study is systematically altered or manipulated. This, as we know, is defined as the independent variable, an element of the study that you as a researcher deliberately and systematically manipulate, change, or select. The effects of the manipulation of the independent variable are examined and measured by the dependent variable. That is, the dependent variable is the observed effect, result, or outcome that is measured in response to a systematic change in the independent variable.

A true experiment provides a strong and powerful test for research that aims to establish a cause-and-effect relationship between independent and dependent variables. In a true experiment, the independent variable is the hypothesized cause, and the dependent variable, the hypothesized effect. A true experiment has high **internal validity** as it provides the best research design for establishing causality between independent and dependent variables. In fact, internal validity, also known as causal validity, is a defining feature of a true experiment that provides a direct test of a causal relationship between independent and dependent variables while controlling for as many extraneous, confounding variables as possible.

For example, Dunn et al. designed an ingenious experiment that allowed them to test their causal hypothesis that prosocial spending promotes happiness. In their randomized experiment, 46 participants rated their happiness in the morning and then were given envelopes that contained either $5 or $20, which they were asked to spend by 5 p.m. that day. Half the participants were assigned to the personal spending condition in which they were instructed to spend the money on a bill, an expense, or a gift for themselves, and half were assigned to the prosocial condition in which they were instructed to spend the money on a gift for someone else or to donate it to charity. Participants were called after 5 p.m. on that same day and again rated their level of happiness. The experimental design thus had two independent variables: amount of money given to participants ($5 vs. $20), which Dunn et al. labeled "windfall," and instructions to spend on either themselves or others, labeled "spending choice" (personal vs. prosocial). Note that an independent variable must, by definition, have at least two **conditions** or **levels**. In this example, the independent variable of windfall has two conditions ($5 vs. $20), as does the independent variable of spending choice (personal vs. prosocial). The dependent variable is the difference in happiness ratings, before and after receiving windfall and spending instructions.

In designing this experiment, Dunn and colleagues were faced with a critical decision: how to assign participants to the different conditions or groups. Recall participants received windfalls of either $5 or $20 with instructions to spend their windfall on either themselves or others. The two independent variables, windfall and spending choice, each with two levels, produced four conditions. This is known as a 2 × 2 (read as two-by-two) experimental design, which we will learn more about in subsequent chapters. As we know from Chapter 1, we want to guard as best we can against any kind of bias in assigning participants to conditions or groups. The best way to do this is to make assignments using chance procedures, such as flipping a coin. This is known as **random assignment**, and it is a critical feature of a true experiment. Random assignment should not be confused with random selection, and *random* here does not mean "haphazard." Rather, random assignment is a simple but indispensable experimental procedure that a researcher uses to help ensure that each participant has the same opportunity to be assigned to any particular condition of a study. For Dunn and colleagues, they randomly assigned participants to one of their four conditions: (1) $5 windfall, spend on self; (2) $5 windfall, spend on others; (3) $20 windfall, spend on self; (4) $20 windfall, spend on others. Using random assignment

provided them with control for biases related to the allocation of participants to one of their four experimental conditions.

Why the term *"true" experiment*? As we will learn in Chapter 8, the term *true experiment* is restricted to those independent variables, such as a placebo and an experimental drug, that can be randomly assigned. In the Dunn et al. study, they used random assignment to devise their two independent variables of windfall and spending choice. In so doing, random assignment helped to ensure that research participants were similar prior to the manipulation of the two independent variables. Thus, the thinking is that any subsequent differences in the dependent variable can be attributed to manipulation of the independent variable. In modern research, the terms *true experiment* and *randomized experiment* are used interchangeably. The modifier, *true*, is, however, misleading, as it seems to connote superiority of a single correct experimental method (Shadish, Cook, & Campbell, 2001).

A randomized true experiment offers the strongest test of a research hypothesis that proposes a cause-and-effect relationship between an independent variable and a dependent variable. Historically, randomized experiments are used to maximize internal validity for studies in which it is important to demonstrate causal relationships. Random assignment is the best method for controlling bias and confounding variables. It allows the researcher to draw the strongest causal inferences free of extraneous assumptions (Wilkinson, 1999). Random assignment makes most other causes and alternative explanations less likely (Shadish et al., 2001).

Randomized experiments are often employed to assess the efficacy of a treatment or intervention. These experiments typically compare three groups: (1) the **treatment group**/condition, also known as the **experimental group**/condition, which receives the treatment or intervention; (2) the **control group**/condition, which receives nothing; and (3) the **placebo** group, which receives an inert substance, or a sham treatment, so that participants in the placebo group think they are receiving the intervention or treatment even though they are not. For example, in many drug studies, the placebo group also serves as the control group or control condition, meaning that it provides a **baseline** to compare to the treatment group. In medical research, when research participants are randomly assigned to receive either the drug under investigation or a placebo, the study is referred to as a **randomized controlled trial (RCT)**. An RCT also often uses a **double-blind** procedure whereby both the researcher and the participants are "blind" to who receives the "real" drug and who receives the placebo. Double-blind and placebo are used to control for the potential confound of unwanted effects of expectations of both participants and researchers that might act as extraneous influences on the results of the study. That is why federal regulations require that the efficacy of new drugs be determined in RCTs: This design makes it highly unlikely that the apparent effect of the drug could really be due to something else. RCTs maximize internal validity.

## Quasi-Experiments

In real life, random assignment is not possible, and so a true experiment is neither practical nor plausible for many scientific questions studied by psychologists. Enter **quasi-experimental** studies (*quasi* means "as if" in Latin) for investigations that aim to examine the effects of an independent variable that cannot be directly manipulated or randomly assigned on a dependent variable. Gender, race, age, ethnicity, socioeconomic status, locale, diagnosis, personality traits, and personal history are just some examples of independent

variables that cannot be directly manipulated by an experimenter. Rather, for these kinds of variables, either an experimenter selects participants who have a particular characteristic, or participants are studied who have been exposed to specified events, such as war or trauma, or who live in certain settings or situations, such as a neighborhood or a geographic region.

Because quasi-experiments by definition lack random assignment, the allocation of participants to specific groups or conditions is done on a nonrandom basis. Participants, for example, will be selected on the basis of certain characteristics, such as a study of people who meet diagnostic criteria for schizophrenia. This nonrandom assignment produces confounds, particularly if the aim of research is to establish a causal relationship. In this instance, a quasi-experiment lacks internal validity, and a randomized experiment is needed. Historically, as evidenced by the prominence of RCTs, scientific research has focused on maximizing internal validity. The priority placed on internal validity is seen in the prominence of RCTs. The idea is that it is more important to know whether a particular intervention works under controlled conditions of a randomized experiment than to know whether it will work across samples, settings, and individuals.

On the other hand, quasi-experiments provide a powerful research design to examine the generalizability of results—that is, whether findings hold across different measures, settings, and persons. In other words, a quasi-experiment places a high priority on **external validity**, namely the generalizability of the research to other settings, populations, and people. Consider the research studies on happiness conducted by Dunn and colleagues. Their randomized experiment provided strong evidence of internal validity in support of the prosocial spending hypothesis of happiness. Yet a key question concerns external validity, namely the generalizability of the finding derived from their randomized experiment that prosocial spending leads to happiness. Do these results generalize to the real world?

Dunn and colleagues thus set out to test the external validity of their idea that prosocial spending leads to happiness by designing a very interesting and realistic field study using a quasi-experiment. They reasoned that if spending money on others is a more effective route to happiness than spending money on oneself, people who receive an economic windfall should experience greater happiness if they spend it on others rather than themselves. Accordingly, Dunn and colleagues recruited 16 employees and examined their happiness before and after they received a profit-sharing bonus from their company. Lo and behold, they found that employees who dedicated more of their bonus to prosocial spending reported greater happiness. Moreover, prosocial spending, not the size of the bonus, predicted happiness.

The Dunn et al. field study of employees is an excellent example of how a quasi-experimental design can address key questions of external validity. The results from their quasi-experiment not only supported the prosocial spending hypothesis but extended their laboratory findings in very important ways. Crucially, it showed that their laboratory findings could be generalized to the real world of business with employees receiving bonuses. A quasi-experiment lacks the scientific rigor of a randomized experiment. Alternative explanations are more probable with a quasi-experiment. Similarly, confounds are more problematic. Ideally, you will try to control for as many variables as possible so that the interpretation of the relationship between the independent and dependent is not unduly confounded by unwanted influences. However, in reality, this is seldom possible, so special statistical techniques may be used to remove the effects of potential confounds. Nonetheless, a quasi-experiment provides a critical complement to a randomized experiment, offering evidence for external validity. We cover quasi-experimental designs in Chapter 11.

## Correlational Research

Correlational (also called relational) research is a widely used study design for examining how two or more variables are associated, linked, or connected. It is a very effective design to address research questions that ask how certain variables are related to each other. The simplest **correlational research** design focuses on the association of just two variables—for example, anxiety and intelligence, happiness and income, or culture and emotion. As we will learn, a **correlation** is a statistic that is computed by a specific formula that indicates how closely two variables are related. A correlational research design can become more complex as in the case of studies that seek to examine how multiple variables might be related, such as an investigation of the relationship of high school grades and College Board scores with subsequent college performance. For these more complex designs, results are often analyzed using a statistical test known as **multiple regression** to examine how two or more variables may be related to an outcome.

For example, using a correlational research design, Dunn et al. conducted a nationally representative survey study to examine the relationship of happiness with personal spending and prosocial spending. To examine this relationship, they used multiple regression techniques to test their hypothesis that prosocial spending, but not personal spending, would be related to happiness. The results of this study provided strong correlational evidence in support for the prosocial hypothesis of happiness. As Dunn and colleagues wrote:

> Although the correlational nature of this design precludes causal inferences, this study provides initial evidence that how people spend their money may be as important for their happiness as how much money they earn—and that spending money on others might represent a more effective route to happiness than spending money on oneself. (p. 1687)

Correlational research is described as nonexperimental. This is because the studied variables are measured but not manipulated, either directly or indirectly. The chief advantage of correlational research is that it allows for the quantitative comparison of variables that cannot be manipulated directly, such as whether two variables are related or correlated. The chief disadvantage of correlational research is that causality cannot be established. We cover correlational research in depth in Chapter 7.

## Sampling and Survey Design

A **survey** is typically composed of a set of questions asking respondents about their activities, opinions, attitudes, or preferences. As a **self-report measure**, a survey provides a relatively inexpensive way to collect a lot of data quickly, and to formulate and test a hypothesis. A survey is, however, limited by what people are capable of reporting accurately. In survey research, it is important to have a representative sample. For Dunn and colleagues, they gathered a nationally representative sample of 632 Americans, 55% of whom were females. For this research, it is also important to have clear, simple, objective self-report survey questions that people can answer easily and on their own. Accordingly, Dunn and colleagues composed several straightforward questions that participants were asked to answer. These self-report measures asked people to rate their happiness and to estimate how much they spent in a typical month on (a) bills and expenses, (b) gifts for themselves, (c) gifts for others, and (d) donations to charity. We cover sampling in Chapter 6 and survey design in Chapter 13.

## RESEARCH IN THE NEWS

### BUILDING PSYCHOLOGICAL KNOWLEDGE WITH SURVEY RESEARCH

"Are Teenagers Replacing Drugs With Smartphones?" reads the headline of an article written by Matt Richtel in *The New York Times* (March 13, 2017). In this piece, Richtel reports on an interesting and curious finding that has been emerging for a decade that American teenagers are growing less likely to try or regularly use drugs, including alcohol. Quoted in the article is Nora Volkow, director of the National Institute on Drug Abuse. She cites recent *survey* evidence showing a steady decline over the past decade in illicit drug use in teens, with the lowest level for 8th, 10th, and 12th graders for the past year. Dr. Volkow wonders whether drug use is declining because of the rapid explosion of cell phones and other interactive media among teens. In this article, Dr. Volkow is quoted as describing these interactive devices "as an alternative reinforcer" to drugs, as she commented that "teens literally get high when playing these games." What do you think? Of course, *The New York Times* is not a scientific journal, and this is not a *peer-reviewed* article. Yet it notes survey data and quotes prominent science researchers. What steps would you take to investigate the scientific claim that drug use is declining among teens?

## Performance-Based Measures

Standardized tests, such as the College Board examination, are perhaps the best example of **performance-based measures**. In psychology, performance-based measures constitute a well-known **psychometric** approach for investigating variables such as intelligence, personality traits, and aptitude. With this approach, test performance is scored and compared to a statistical average derived from a normative or standardized sample taken from a wider population.

For example, a well-known psychometric test of intelligence is the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV; Wechsler, 2008a, 2008b). This test provides a measure of a variety of cognitive abilities, such as vocabulary, mental arithmetic, spatial reasoning, and using blocks to construct designs, performance on all of which is summarized in the form of an intelligence quotient (IQ). IQ is computed by comparing a person's test score to that of an average score of a normative group of peers of similar ages.

The advantage of the psychometric approach rests largely on the extensive reliability and validity studies that are performed in the development and construction of an instrument, such as the WAIS-IV. An important disadvantage, however, is that psychometric measures are often criticized for being **culturally biased** in their construction and selection of test items and tasks. An additional important consideration is the extent to which the normative or standardization sample for a psychometric test is truly representative of the diversity of the general population, or an appropriate benchmark for an individual test taker of an ethnic or racial minority. Performance-based measures are covered in Chapter 5.

## Qualitative Research

In many instances, researchers may seek to study questions for which numerical or empirical answers may not provide the most complete answers, such as questions of values and meanings people attach to human behavior and beliefs. For example, how and why is it that people from Western cultures tend to value independence and individualism in contrast to the high value people of Eastern cultures give to interdependence and collectivism? An even deeper question might be to address the fundamental nature of such categories as individualism and collectivism. For these questions of how and why, researchers often use what are referred to as qualitative methods, such as participant observation, intensive interviewing, or focus groups. These techniques are intended to *go beyond numbers* in order to study and understand phenomena in terms of the meanings people attach to them, thereby preserving and capturing the complexity and diversity of human behavior.

Often used in qualitative research, a **naturalistic observation** design studies people in their natural settings so that their behaviors and words can be put into their proper context. Such descriptive study of people is also sometimes referred to as **ethnography**. Here *observation*, it is important to emphasize, does not mean the casual "seeing" of everyday life that leads to haphazard impressions. To the contrary, for this research methodology to be effective, observation must be controlled or *systematic*, which means that it must be conducted carefully, with precise description that allows for consistent or reliable cataloging of data and the orderly classification and analysis of that information (Adler & Adler, 1994).

While qualitative researchers use naturalistic observation methods that tend to avoid predetermining categories of action that can be precisely measured, they, like their quantitative counterparts, make sure their studies yield reliable and valid data. In short, the aim of qualitative research is to understand context—the what, how, when, and where of an event or an action. It yields data regarding meanings, concepts, definitions, characteristics, metaphors, symbols, and descriptions of events or actions (Berg, 2004). As such, it is ideally suited for an inductive research approach.

Consider a simple but interesting study by Hussain and Griffiths in the Psychology Division at Nottingham Trent University. Hussain and Griffiths (2009) were interested in the attitudes, experiences, and feelings of online gamers. Although massively multiplayer online role-playing games (MMORPGs) have become very popular and there have been some studies of these online gamers, there has been little qualitative research exploring gamers' accounts of their own activities and attitudes. To start to fill this gap, Hussain and Griffiths recruited 71 online gamers through posts on online gaming forums and in *World of Warcraft* games. The researchers "interviewed" these 71 participants online, either through online chat or by email (see Exhibit 2.3). Most of the interview questions were open-ended. The researchers explained why they used this approach:

> The unstructured nature of the interviews allowed gamers to develop their own narrative by exploring their experiences of MMORPGs. The researcher allowed gamers to speak for themselves (i.e., the emergent themes were participant led rather than researcher led). This allowed gamers to take control of the interview process and prevented researchers' subjective bias entering the analytic stage. (p. 748)

The researchers read through the interviews and identified the main themes that were expressed in them. For example, many comments had to do with the psychosocial impact

of online gaming. Although the "vast majority" of the gamers highlighted positive effects of the activity, some pointed to potentially adverse effects:

> I've lost my IRL [in real life] friends because I couldn't find the time to be with them; I quit school. Whenever someone asks me to do something on the weekends, I always think "ooh, but we're raiding, I really shouldn't go out," and that's a way of thinking which I really dislike. (p. 750)

The researchers concluded by noting how gamers use gaming to alleviate negative feelings but also how they may experience personal problems due to the online gaming. They raised questions about the difference between socializing online and offline, and they focused attention on the problem of addiction. Although Hussain and Griffiths did not attempt to develop their own theory of online gaming, the patterns they found and the questions that emerged from their research provide a foundation that others could use in developing such a theory. The researchers urged more generalizable research about online gamers using quantitative methods, and they recommended policies that could reduce the likelihood of problems due to online gaming. Chapter 14 covers qualitative research in detail.

## Literature Review

Literature reviews represent yet another tool for research. Indeed, all psychological research presented in peer-reviewed journals generally begins with a literature review. Some psychology journals are exclusively devoted to publishing literature reviews of particular

**EXHIBIT 2.3 ■ Interviewing Online Gamers**

*Source:* ©iStockPhoto.com/Micko1986.

topics of interest. Consider, for example, what the prestigious *Psychological Bulletin* seeks to publish, as described on its web page:

> Integrative reviews or research syntheses focus on empirical studies and seek to summarize past research by drawing overall conclusions from many separate investigations that address related or identical hypotheses. A research synthesis typically presents the authors' assessments of (a) the state of knowledge concerning the relations of interest; (b) critical assessments of the strengths and weaknesses in past research; and (c) important issues that research has left unresolved, thereby directing future research so it can yield a maximum amount of new information. (www.apa.org/pubs/journals/bul)

As you can see from this description, literature reviews serve an extremely important function in research. A sound literature review creates and evaluates a body of knowledge that is drawn from a synthesis of empirical findings across independent studies. We cover literature reviews in Chapter 3.

## EVALUATING AND CRITIQUING RESEARCH

We have now learned the diversity of methods that are available to a researcher. Each has its pros and cons. The advantages and disadvantages will depend on many factors, not least of which, as we have seen, relates to the research question under investigation, the theoretical model being examined, and the hypothesis being tested. In reality, a researcher would be wise to make use of these different methods in pursuing a theoretical and practical understanding of a topic. Indeed, multiple methods and techniques that produce converging findings offer some of the strongest evidence in support of a research hypothesis and psychological theory. As we will learn, converging findings adds to both reliability and validity of the research.

We now turn our sights to thinking about how to make scientific sense of these various research designs with respect to how they are applied in psychological studies. What approach might you take in evaluating and critiquing research? What features or benchmarks might you consider in weighing the pros and cons of a research study?

### Scientific Skepticism

**Skepticism** has long been a defining feature of a researcher's mindset. Indeed, skepticism no doubt inspired Galileo when in 1612 he disproved the geocentric theory of the universe and proved the heliocentric theory of the solar system. Under the principle of scientific skepticism, a researcher doubts, questions, and withholds judgment until sufficient scientific evidence has been established. Scientific evidence is the final arbiter of the validity of a hypothesis, psychological theory, or research study. Today, in clinical psychology, the strong emphasis on *evidence-based* treatments is yet another reminder of the value of a skeptical mindset.

Skepticism, with its philosophical roots in **logical positivism** and **logical empiricism**, requires critical thinking using the principles embedded in the scientific method covered in Chapter 1. In scientific research, there is no authority figure to establish knowledge. Rather,

knowledge is accumulated and gained through studies grounded in the scientific method. A research study is examined not only on its own scientific merit, but also in relation to other similar investigations. Indeed, a scientific skeptic does both, by critically assessing the theory and design of a single study and by asking how it adds to the cumulative nature of knowledge. By cumulative nature of scientific knowledge, we mean that as more findings are discovered, the more we progressively come to know about the research topic under investigation. The creation of psychological knowledge is thus a cumulative process.

## Peer Review

Peer review plays a critical role in how psychology and science builds knowledge. Described as "a gatekeeper, the final arbiter of what is valued in academia" (Marsh, Jayasinghe, & Bond, 2008, p. 160), peer review is used to evaluate journal submissions and grant proposals in psychology and other scientific disciplines. It is seen as a quality control system, the so-called seal of approval in establishing the trustworthiness of research (Marsh et al., 2008). **Peer review**, as the cornerstone of psychological science, is the process by which a research study is assessed for quality before it is published. Also known as **refereeing**, peer review works as follows: Scientists conduct a study or review literature on a topic and write up their results in a research article format, which they then submit to a scientific journal. The journal editor in turn sends the research article out for review to usually two or more referees who are charged with evaluating the scientific quality of the article. The referees then submit their reviews to the editor who makes the final decision on publication.

To illustrate, let's consider the study by Dunn and colleagues published in the journal *Science*. As we learned, in this article, Dunn and colleagues reported the results of three different studies that they conducted addressing the question of happiness. As we learned, the results of each of these studies—correlational, quasi-experimental, and randomized experimental—provided converging evidence in support of their prosocial happiness hypothesis. Upon submission to *Science*, the editor sent the Dunn et al. manuscript out for review to referees. Based on the referees' reviews, the editor decided to publish the Dunn et al. study, with the article subsequently appearing in the journal *Science*.

In scientific research, we focus *only* on papers published in peer-reviewed journals. There are many peer-reviewed journals devoted to psychological research, and these and all other scientific journals are ranked using a metric known as the **impact factor** (Garfield, 1999). The impact factor is simply the average number of times a journal's articles have been cited by other researchers over a specified period of time, often two years (Larivere et al., 2016). A higher average number of citations is indicative of greater impact for the journal. This is often interpreted as an indicator of the quality of the journal and its articles. The impact factor for *Science*, where Dunn and colleagues published their paper, ranks among the highest for peer-reviewed journals.

Both impact factor and peer review are important global benchmarks for evaluating research. Understanding them will help you as you learn about research throughout this book. Each, however, has been subjected to important criticism. For example, the impact factor, while very useful for ranking peer-reviewed journals, is considered a poor indicator for assessing the importance of a particular article. This is because a given journal's impact can be unduly influenced by a small fraction of highly cited articles. The vast majority of papers appearing in the same journal may receive very few, if any, citations. This results in a skewed distribution of citations, rendering the average number of citations for a given

journal as misleading (Larivere et al., 2016). In other words, a research study published in a journal with a high impact factor is not necessarily more likely to be cited.

Evaluating the impact of research plays an important role in the peer review process. But an important question for us is what research principles and criteria underlie and guide peer review and how these might be applied in evaluating and critiquing research. Questions about how to evaluate research are essentially questions about reliability, validity, generalizability, and biases. As Marsh and colleagues (2008) wrote, these traditional psychological criteria have too often been lost in the peer review process that is so crucial for evaluating research and building psychological knowledge. Now let us look at how these psychological criteria may be applied in evaluating what is good in research!

## Reliability of Measurements

A general assumption of research is that measurement scales used in a study are reliable, meaning that they will yield similar scores when administered in the same conditions (Diener, Inglehart, & Tay, 2012). You as a reviewer of a research article can and should check this assumption by, for example, looking to see if the paper reports reliability for any scales used in the study (Wilkinson, 1999). In many instances, because reliability of measurement is so central to sound research design, there is often an extensive body of literature investigating the topic. For example, consider the measurement scale of happiness used in the Dunn et al. study. Diener et al. (2012) extensively reviewed studies examining the reliability of this measure of happiness, also known as the Satisfaction With Life Scale (SWLS). Their review indicated that the SWLS is stable, yet sensitive to changes over time. They concluded that the evidence indicated the SWLS has both high internal consistency and high stability, and thus provides a reliable measure of happiness.

## Validity and Generalizability

Let us begin with a simple question. Why is it so important to use different methods and research designs to study a psychological question? Put simply, we want to see if there is a **convergence** of results across different methods and designs. As we saw in the Dunn et al. research, three different studies produced similar results. This convergence of results offered strong support for their prosocial spending hypothesis of happiness. It provided strong evidence of the validity of their research and psychological theory of happiness. Recall their randomized experiment offered strong support for internal validity whereas their correlational study and quasi-experiment offered strong support for external validity. Not surprisingly, then, their paper fared well enough in peer review to be published in the prestigious journal *Science*. That their research provided evidence in support of these two critical elements of validity—internal and external—no doubt impressed the referees and editor.

As we have emphasized, each form of research has its own strengths and weaknesses, so using two or more different designs to study the same psychological question will go a long way in demonstrating the validity of your research. However, some researchers may be biased to one type of research design over the other—for example, believing that a randomized experiment offers more valid information than a survey study. This can lead to biases in evaluating and critiquing research. In fact, the peer review process in psychology has been criticized for its overemphasis of internal as opposed to external validity (Sue, 1999). This is an important criticism as we think about evaluating research. This means

that scientific psychological knowledge may be biased toward information and empirical evidence derived from studies with high internal validity but low external validity. Sue (1999) has described this emphasis on internal validity at the expense of external validity as the "selective enforcement of scientific principles" (p. 1072).

What is the most serious implication of favoring studies of high internal but low external validity? Low external validity means that research findings have very limited generalizability to a wider population beyond the sample studied. This, of course, represents a serious limitation, as a central aim of psychological science is the development and accumulation of generalizable knowledge. Ideally, of course, we would aim for designing a program of research that would place equal emphasis on external and internal validity. Indeed, this was the case for Dunn et al., who used empirical evidence gathered from both a nationally representative survey study and a quasi-experiment of employees to demonstrate external validity, and data from a randomized experiment to demonstrate internal validity.

Nonetheless, we, as peer reviewers, may still have serious concerns about the external validity of the Dunn et al. study. So too apparently did these researchers, who over the next several years following publication of their *Science* article designed a number of studies addressing the external validity of their finding of the positive influence of prosocial spending on happiness. Consider the 2013 paper authored by Aknin and colleagues, "Prosocial Spending and Well-Being: Cross-Cultural Evidence for a Psychological Universal," published in the high-impact *Journal of Personality and Social Psychology*. In this paper, the researchers asked the interesting question as to whether the positive relationship between prosocial spending and happiness represents a **psychological universal**, meaning a quality shared by humans everywhere but expressed in varying degree according to cultural context (Norenzayan & Heine, 2005).

Crucially, these researchers found strong support for the universality of the prosocial spending link to happiness, and most impressively, they demonstrated this relationship in both correlational and experimental studies. First, their survey data collected from 136 countries found prosocial spending to be positively associated with happiness around the world, in both poor and rich countries. Second, their experimental data showed that asking people to recall a past instance of prosocial spending had a causal impact on their happiness across rich and poor countries (Canada, Uganda, and India). Finally, in yet another experiment, Aknin et al. (2013) randomly assigned participants in Canada and South Africa "to buy a goody bag filled with treats for either themselves (personal spending) or a sick child at a local hospital (prosocial spending)." (p. 644). Replicating the Dunn et al. finding, they found that the participants randomly assigned to the prosocial spending condition reported higher levels of happiness than did participants assigned to the personal spending condition.

In another study, Aknin, Hamlin, and Dunn (2012) tested their prosocial spending hypothesis of happiness in toddlers. Here the researchers reasoned that if the capacity to derive happiness from charity is a universal feature of human psychology, then even very young children might show this emotional benefit from giving to others. Accordingly, they composed a sample of toddlers, just under the age of two, and presented them with a bowl of appealing treats (goldfish crackers) along with a puppet. The experiment included two key conditions for testing whether sharing leads to happiness even in children. In one condition, the experimenter "found" a treat (not from the toddler's bowl) and asked the toddler to give this treat to the puppet. In the other condition, the experimenter asked the toddler to share a treat from her or his bowl with the puppet. Thus, the only difference was in one condition

toddlers gave the "found" treat to the puppet and in the other condition they gave away their own treat to the puppet. Their results indicated that toddlers were happier giving away their own treat than when giving away the found treat.

## Multicultural Analysis

Cross-cultural psychology is one of most influential schools of psychology today. Its focus is on examining and understanding similarities and differences across cultures (Matsumoto & van de Vijver, 2011). It studies how individuals and cultural variables influence human psychology. Its numerous insights about cross-cultural comparisons have contributed to a richer and more nuanced understanding of the mind and human behavior.

Cross-cultural psychology offers many important lessons for us in evaluating the scientific merit of research studies. Cross-cultural psychology often uses quasi-experimental designs, for which we know random assignment is not possible. As Matsumoto and van de Vijver (2011) noted, "researchers cannot randomly assign an individual to a culture" (p. 2). This, of course, makes interpreting results of similarities and differences much more challenging for cross-cultural studies than for randomized experiments. Alternative or rival explanations for findings of cultural differences are much more difficult to rule out with quasi-experimental designs. But perhaps because of these challenges, cross-cultural studies illustrate and reinforce two important key principles for designing reliable and valid research.

First is the importance of careful description of a study sample. A study sample should be inclusive and diverse, and these cultural variables should be carefully measured and used to stratify the sample into meaningful groups. This is so important because the strength of a psychological research report depends largely on the extent to which findings can be generalized from a study sample to a broader population. Using carefully defined samples rich in diversity yet divided using measurable cultural variables will allow the researcher to specify explicitly the populations to which the findings can be generalized. As Sue (1999) aptly wrote, "This is the external validity issue. If we want to draw conclusions about human beings, we must study human beings in all their diversity and not particular samples" (p. 1076).

## STAT CORNER

### SAMPLE SIZE MATTERS

The *law of large numbers* states that the larger the sample size, the more reliable and valid the sample statistics that are used to estimate population parameters. This is a fundamental principle of statistics. Consider this example taken from Nisbett, Fong, Lehman, and Cheng (1987), in which research participants are asked to explain this simple observation: Why, after the first two weeks of the Major League Baseball season, is the highest batting average .450, but no player in the history of the game spanning a century has ever had as high an average as that by the end of the season? The answer to the question is instructive on two grounds. First, two weeks provides a relatively small sample of a player's batting ability. Second, extreme values, such as a .450 batting average, are more likely in a small sample than in a large sample (see also Tversky & Kahneman, 1974). So, as you learn to evaluate research, always keep in mind the size of a study sample, as the law of large numbers is relatively straightforward to understand yet too often overlooked or ignored.

A second important lesson pertains to the external validity of psychological measurements. For example, we know that the norms for a psychometric test, widely used in research, are derived from a standardization sample. However, a particular psychometric test may not be valid if, for example, an examinee's cultural ethnicity was not included in the original standardization sample. Similarly, there is often the assumption that study variables, such as happiness, are similar across cultures. This, however, needs to be tested, as the establishment of **measurement invariance** of key study variables is essential for valid cross-cultural comparisons. In fact, Bieda et al. (2017) found that in a large and diverse sample that included German ($n$ = 4,453), Russian ($n$ = 3,806), and Chinese ($n$ = 12,524) university students, scales of happiness showed strong cross-cultural measurement invariance. These data offered evidence that the nature of happiness may be similar across cultures. However, whether such cross-cultural invariance would hold for other psychological variables is unknown and requires empirical study similar to that conducted for happiness by Bieda et al. (2017).

## CONCLUSION

Research comes in many different forms, yet all share the same goal—the accumulation of reliable and valid psychological knowledge. The two major research approaches—nonexperimental and experimental—offer different yet complementary methods to investigate a question, to evaluate a theory, and to test a hypothesis. Both approaches can call upon a toolbox of methods and study designs. Here we used a set of published studies examining the prosocial hypothesis of happiness to highlight three distinct kinds of methods and research designs—correlational, quasi-experimental, and a randomized experiment. Each, we learned, provides valuable information, with high external but low internal validity for evidence derived from survey and quasi-experimental studies in comparison to high internal but low external validity for randomized experimental designs. All research designs require reliable measures, which can be evaluated statistically (see Chapter 5). Science focuses only on peer-reviewed research, which is evaluated using principles of reliability and validity. In fact, cross-cultural research reinforces the critical importance of reliability and validity in evaluating evidence and in optimizing the generalizability of psychological knowledge.

## KEY TERMS

Baseline
Condition
Control group
Convergence
Correlation
Correlational research
Culturally biased
Deductive research
Double-blind
Empirical generalization
Ethnography
Experiment
Experimental
Experimental group
External validity
Generalization
Impact factor
Inductive research
Internal validity
Levels
Measurement invariance
Multiple regression
Naturalistic observation
Nonexperimental
Peer review
Performance-based measures
Placebo

Psychological universal
Psychometric
Quasi-experimental
Random assignment
Randomized controlled trial (RCT)
Refereeing
Research circle
Self-report measures
Skepticism
Survey
Treatment group

## ACTIVITY QUESTIONS

1. In "Research in the News," we presented an article appearing in *The New York Times* (March 13, 2017), "Are Teenagers Replacing Drugs With Smartphones?" by Matt Richtel. This is a piece of journalism that is not peer reviewed for its scientific merit but nonetheless reports empirical evidence from survey data showing robust decline in illicit drug use in teens for the past decade. Imagine that you went to the primary source—that is, the peer-reviewed publications that reported on the actual survey data. Describe and explain how you would scientifically review and critique these survey data.
2. Richtel's *New York Times* piece reports on two interesting findings or trends—decline in teen illicit drug use and explosion in teen smartphone use. You want to investigate the association, if any, between these two findings. Describe your research approach and specific study design. What are the pros and cons of your research design? What kind of inferences might you be able to draw if you found declining illicit drug use to be associated with smartphone use?
3. Describe an experimental approach that would test the relationship of teen illicit drug use and smartphone use in teens. State your hypothesis and describe your study design. What alternative explanations and confounds would you consider? How might you effectively rule out alternative explanations?

## REVIEW QUESTIONS

1. Compare and contrast quasi-experimental and randomized experimental research designs. In your analysis, explain the pros and cons of each design with respect to external validity and internal validity.
2. Describe and explain how experimental and nonexperimental approaches were used to provide converging evidence in support for the prosocial spending hypothesis of happiness.
3. Compare and contrast the complementary roles of reliability and validity in evaluating research.
4. Describe peer review and its function in scientific research. What scientific principles guide peer review?
5. How would you design cross-cultural research aimed to maximize the generalizability of psychological knowledge? Be sure to address the importance of sampling and reliability of measures in designing culturally sensitive research.

## CHAPTER 2 TESTING BEFORE LEARNING ANSWERS

**ANSWER KEY:** 1. d; 2. c; 3. a; 4. b; 5. b; 6. c

**SAGE edge** provides a personalized approach to help students accomplish their coursework goals in an easy-to-use learning environment. The site includes flashcards for key term practice, learning objectives to reinforce key materials, along with open access media for concept exploration. Visit the site at **https://edge.sagepub.com/nestor3e.**