iStock.com/Kenneth Cheung

2 The Science of Psychology

Research Methods and Statistics

Is social media ruining the world?

If it is, how would we know? Psychology's methods and statistics empower us to find out. We (your authors) use social media and are willing to bet that many of you taking this course also use at least one form of it. But is social media making our lives better, or worse?

There must be some benefits, or we wouldn't spend so much time posting and scrolling through the apps. In fact, research shows that social media helps people keep up with trends, read the news, share opinions and photos with friends and family, find communities of people with similar interests, and indulge in hours of distraction when we just don't feel like getting our homework done (or grading papers; e.g., Khan et al., 2014; Pinter et al., 2021; Uhls et al., 2017).

But many people worry about social media's potential negative side effects, ranging from car accidents, cyberbullying, gambling, and even promoting terrorism (Hashash et al., 2019; Juergensmeyer, 2017; Khan et al., 2014; Pinter et al., 2021; Uhls et al., 2017; Weimann, 2016). People with high levels of anxiety also worry about how they use social media, sometimes worrying about if they're addicted (Bhandarkar et al., 2021; Bowden-Green et al., 2021; Praveen et al., 2020; Zhong et al., 2021). This list of anxieties could be a lot longer, but you get the idea. Research indicates that social media can have both positive and negative effects. But it's unclear whether using social media causes these negative outcomes or if these variables simply are related for other reasons.

How do scientists know all this information—and how do psychological studies work in general? You can form your own opinion about social media, or anything else, with more ease and confidence when you understand research on both sides of a debate. Even if you never do research yourself, being a good consumer of research will help you know that your opinions are based on solid, scientific evidence. It will also help you see how hard psychologists work to make sure psychological ideas are tested with real data from real people, all over the world.

After reading this chapter, you will get answers to several questions you've been curious about:

Have You Ever Wondered?

2.1 How is psychology a science?

2.2 How do psychologists design studies?

2.3 What are correlations?

2.4 How are the results of experiments analyzed?

2.5 How can I tell if a study is done well?

2.6 What is the open science movement?

2.7 How do ethics guide psychological research?

Learning Objectives

LO 2.1 Explain how the scientific method is used in psychological science.

LO 2.2 Compare and contrast research methodologies.

LO 2.3 Explain how we show relationships between variables and why correlation doesn't imply causation.

LO 2.4 Compare and contrast *t*-tests and ANOVAs.

LO 2.5 Summarize how studies can be analyzed for quality.

LO 2.6 Explain how the open science movement is addressing the replication crisis in psychology.

LO 2.7 Describe ethical considerations for research with humans and animals.

ELEMENTS OF THE SCIENTIFIC METHOD

>> LO 2.1 Explain how the scientific method is used in psychological science.

? Have You Ever Wondered?
2.1 How is psychology a science?

How is psychology a science?

Your high school memories of science and scientists may include learning about rocks and minerals, perhaps using a Bunsen burner or a microscope, and applying equations that describe force and acceleration. You might be familiar with famous scientists like the pioneering chemist Marie Curie, or maybe Katherine Johnson, the lunar landing mathematician whose life was dramatized in the film *Hidden Figures*. Or maybe a documentary from astrophysicist and science communicator Neil DeGrasse Tyson stirred your interest in the cosmos. How well does psychology fit with these other sciences?

(a) Bettmann/Contributor/Bettmann/ via Getty Images

(b) NICHOLAS KAMM/AFP/via Getty Images

(c) Amanda Edwards/Getty Images Entertainment/via Getty Images

(a) Two-time Nobel Prize–winning physicist and chemist Marie Curie said, "Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."

(b) Katherine Johnson, lunar landing mathematician, noted, "I tried to go to the root of the question."

(c) Neil DeGrasse Tyson, astrophysicist and science communicator, explained, "Knowing how to think empowers you far beyond knowing what to think."

Katherine Johnson had to understand the changing gravitational mathematics of outer space to calculate the lunar landing. The scientific approach to physics and math is objective—and gravity never gets bored, misunderstands instructions, or lies on surveys about personal needs. In contrast, the science of psychology is challenging because understanding human and animal thought and behavior requires essential considerations for ethics. In addition, many abstract ideas—like love and prejudice—can be defined and measured in a wide variety of ways. Some psychologists in the past didn't use scientific methods by today's standards.

For example, Sigmund Freud was an original thinker and highly influential. But while Freud's ideas "contain most interesting psychological suggestions," observed the science philosopher Karl Popper, they were not presented "in a testable form" (Popper, 1963, pp. 33–39). Modern psychologists pride themselves on doing things differently. Theories need to be stringently tested multiple times before we can start to feel confident that we understand something. Doing scientific studies with humans and animals is, in a lot of ways, more challenging than mixing acids together or measuring how quickly an apple falls from a tree. People are complicated, and any research must be ethical. How does all of that happen?

The Cycle of Science

Science never ends—it's a cycle that constantly evolves.

Scientific method: A series of objective steps for empirically testing an idea.

Any discipline or field of study—physics, chemistry, biology, psychology—can only claim to be a science if those who practice it use the **scientific method**. It's a general approach to forming and testing ideas using an objective framework. The scientific method uses evidence to determine whether an idea has merit, needs to be changed and refined, or should just be discarded altogether. It's a framework that proposes ideas tested with methods that are *reliable*, *observable*, *testable*, and *valid*. (Memory helper: Think of them as the ROTV requirements.)

The scientific method relies on "ROTV" evidence to evaluate how a hypothesis needs to be changed, refined, or rejected in service of the next, better question. It's a series of steps, displayed in Figure 2.1.

Hypothesis: A specific statement about the expected outcome of a study.

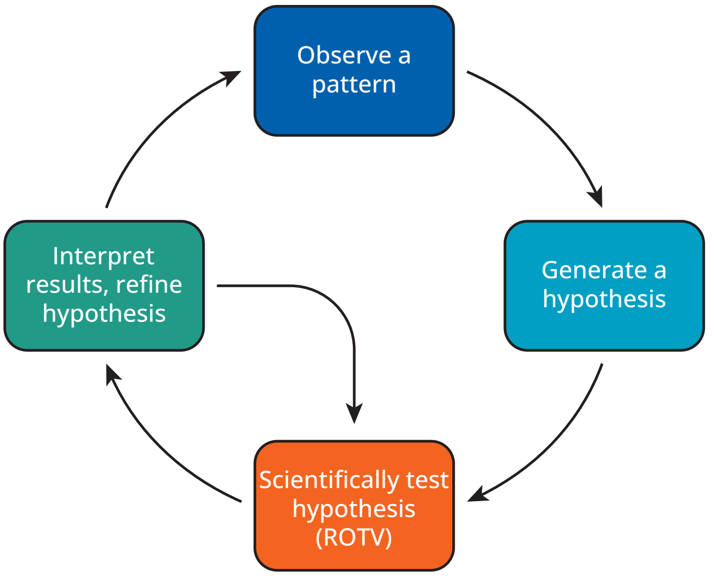
The scientific method usually starts when people perceive some pattern and become curious. That curiosity leads to a **hypothesis**, a specific statement about the study's expected outcome. For example, imagine a professor notices that some students seem to be checking social media during class instead of taking notes—and that those same students are struggling with their grade (a pattern has been observed). The professor forms the hypothesis that the more time students spend on social media in class, the worse their grades are.

Once a hypothesis is generated, the scientist can choose some way to use the ROTV requirements to test the hypothesis. There are dozens of different ways that might happen, and many of those options will be explained in this chapter. For now, imagine the professor used a survey asking students to estimate how many minutes per hour in class they spend looking at social media (instead of paying attention and taking notes). The professor might then compare each person's estimate with their grade, testing for the hypothesized pattern.

FIGURE 2.1

The Scientific Method

The scientific method starts by noticing interesting patterns and generating hypotheses regarding those patterns. Then, evidence is gathered that either supports or refutes the hypothesis. Those results help us refine hypotheses and keep testing them as we learn more.



However, even if the data support the hypothesis, the professor won't fully believe it because the class may be a unique case, and there are alternative explanations. If the hypothesis seems right, the professor might want to double- and triple-check with other classes to see if the same pattern holds up. From there, they could keep going by adding to the hypothesis, making it more complicated or focused, or testing it in a variety of ways. That would inspire more research studies—and the process keeps going.

The Path to Precision: Defining and Measuring Constructs

You can't take an X-ray of self-esteem or anxiety.

Psychological concepts are often abstract and invisible; these features make the path to scientific precision more challenging. On the other hand, gravity is also invisible—but that never stopped Katherine Johnson from measuring its effects. So, imagine this hypothesis: Greater use of social media is associated with greater anxiety. That's a good hypothesis, and many studies have found evidence supporting it (e.g., Dobrean & Păsărelu, 2016; Keles et al., 2020; Vannucci et al., 2017). But what exactly is meant by "anxiety"? For that matter, what counts as "use of social media"?

Science requires precision. Hypotheses identify variables of interest, and in psychology, these variables might be concepts like anxiety, personality, intelligence, love, conformity, and so on. Abstract concepts or variables of interest like these are called **constructs**, which is a useful term because it reminds us that we have built a concept (like a construction crew) into something we can measure using the ROTV requirements.

Construct: An abstract concept or variable within a research study.

The first step in building the construct is performing **operationalization**, which is specifying how we are going to define and measure a particular variable.

Operationalization: Specifying how a construct will be defined and measured in a given study.

Researchers might operationalize anxiety differently:

- One researcher might measure "anxiety" through physiological measures like heart rate and sweaty palms after asking a person to imagine being rejected by friends.
- Another might use surveys or checklists of anxiety symptoms, such as nervousness or sleep difficulties.
- A third approach might test how well people can concentrate under time pressure by asking them to remember lists of random letters or words.

If you think these multiple ways of testing anxiety complicate things, then you are correct. In a paper that reviewed the connections between social media use and anxiety, researchers noted that these different methods made it more complicated and difficult to know what to believe (Keles et al., 2020).

However, it's not necessarily a bad thing to rely on various definitions and measurements. If different measurement approaches find similar connections between social media use and anxiety, then we are more confident that the connection is real. It's like the old example of people who can't see—they can only feel—trying to describe an elephant. One is holding the tail, another the trunk, and a third an enormous foot; they all get different impressions of the same animal. They have good reasons to disagree, but each of their measurements or observations have merit. Only by combining their information do we form an accurate overall picture.

When you start doing your own studies, one of your first jobs is to operationalize your variables. Once you have established the hypothesis, the variables, and how to define and measure them, you can select a research methodology.

Practice what you learn in **Knowledge Check 2.1**

TYPES OF RESEARCH

Have You Ever Wondered?
2.2 How do psychologists design studies?

>> LO 2.2 Compare and contrast research methodologies.

How do psychologists set up research studies?

You will learn about the many creative ways psychologists test hypotheses during this course. This chapter focuses on the strengths and weaknesses of four common methods: archival studies, naturalistic observation, surveys, and experiments. There will be examples of each in this chapter, but as you progress through the course, consider the methodology chosen by the researchers and whether it was the best choice—and why.

The options covered here are summarized for you in Table 2.1. Note that in addition to these specific choices, researchers also have to decide whether to do a quantitative or qualitative design. Again, there are advantages and disadvantages to each choice. *Qualitative* research is usually (but not always) done with surveys or interviews, and the results are in nonnumerical form. The researcher will identify patterns or themes across people's answers. A big advantage of qualitative designs is that it honors the personal voices and experiences of the participants; it also provides deeper insights into individual lives. But a disadvantage is that the results generally can't be analyzed with most statistics. In contrast, *quantitative* methods are any for which the results are in numerical form—and therefore, patterns can emerge through mathematical, statistical analyses.

Archival Studies

The first option might be the most convenient way to test hypotheses.

Archival studies: Research using materials originally created for some other purpose, like police records or social media posts.

Archival studies analyze materials that were originally gathered, produced, or published for some other purpose. Patterns can be observed in newspapers, census data, police reports, college transcripts, and—increasingly—on the Internet. Data for archival studies can be found anywhere that has stored information.

TABLE 2.1

Pros and Cons of Methodology Options

Archival Studies	Pro: Data are publicly available and usually inexpensive
	Con: Data are limited to what already exists
Naturalistic Observation	Pro: Behavior is authentic and honest
	Con: Reactivity and uncontrolled environment
Surveys	Pro: Large samples can be gathered relatively quickly
	Con: Answers may not be honest; conclusions may be only correlational in nature
Experiments	Pro: Random assignment to condition allows for causal conclusions
	Con: Relatively expensive, time-consuming, and/or slow

Every methodological option has advantages and disadvantages.

Studies on social media commonly use archival data because social media users produce so much material to be analyzed! One study investigated if there was a difference between what men and women posted to Facebook (Park et al., 2016). Do the topics, and even the words, used on status updates differ? They found that girls and women were more likely to post things related to social relationships (such as their close friends and family) and emotions. In contrast, boys and men were more likely to focus on specific activities (like sports or jobs) and objects (like computers).

Another study analyzed posts from college students during a presidential election (Carlisle & Patton, 2013). The researchers examined each profile for 50 different characteristics, such as political opinions or support for a certain candidate. Not surprisingly, political posts increased as the election got closer. A more interesting finding was that having more online "friends" predicted fewer political postings. Keep in mind that even when a study answers one question, it may create others such as: Why might having more friends discourage political posts? Every study can inspire more.

There are pros and cons to archival studies. One big advantage is that the data sources already exist and are ready to be analyzed (often, for free!). But a major disadvantage is that the information available is limited, and its availability can't be controlled. So, researchers often also consider other methodology options.

Naturalistic Observation

Have you ever sat and done some "people watching" over a cup of coffee?

Naturalistic observation is a bit like that, but if you also systematically recorded the behaviors you witnessed. This technique observes people in their natural environments—where they would have been anyway. We can observe littering in parking lots, how shoppers navigate the aisles in grocery stores, and how much time men versus women need to complete ATM transactions. Anywhere public will do. Researchers choose a particular spot where they can test a hypothesis or research question in an authentic way, because (ideally) the people being watched have no idea they're part of the study at all.

Naturalistic observation: Watching and recording people's behaviors where they would have happened anyway, but for research purposes.

One observational study offered some archival value after the COVID-19 pandemic. Researchers wanted to observe hand hygiene among medical students (Kwok et al., 2015). To observe their behavior, the scientists videotaped 26 medical students in a classroom and watched how often they touched their faces, which was an average of 23 times per hour. Of those times, 44% of the touches were to mucous membrane areas of the face: eyes, nose, and mouth. Observation is important here, because it is likely many of the students had no idea how much face-touching they were doing—so asking them to report it in a survey wouldn't have produced accurate results.

A clever example of naturalistic observation and social media studied how romantic couples act when one of them starts "phubbing" the other in public (Franz, 2014). What's phubbing? It's "phone snubbing"—when the person you're with is more interested in their phone than they are in talking and looking at you. Rude, right? To study this, the researcher walked through local farmers markets looking for people who appeared to be couples—then waited until one or the other started using their phone for (apparently) checking or posting to social media.

How would the other person react? Three common responses were observed:

(1) Indifference: Ignoring what was happening and looking away.

(2) Collaboration: Participating by also looking at the device and talking about whatever was on the screen.

(3) Assertiveness: This included purposely walking away to show annoyance or even taking the device away from the other person.

Women were more likely to collaborate; men were more likely to show indifference or be assertive about not liking what was happening. Note a disadvantage to this study: The people who shop at a farmers market might act differently in that situation than in others (say, at home). And the results might vary among people with different backgrounds, experiences, cultures, ages, and so on. Maybe behavior at farmers markets in California are different from those in Ohio. Researchers can't control all of these variables, so it's important to know the limitations.

There are both advantages and disadvantage to naturalistic observation. We get to see how people really act in the "real world." Their behaviors are authentic and honest because they don't realize they're being watched. In addition to the problems noted in the previous paragraph, there are also potential ethical problems because participants have not given their permission to being observed for research purposes.

Here's another problem: What if the people being observed *do* suspect that they are being observed? **Reactivity** occurs when people change their behavior when they believe they're being watched—so the behavior is no longer authentic. How can researchers get around the problem of reactivity? **Participant observation** is one solution; it's when researchers, like spies, go "undercover" pretending to be part of the natural environment. In the phubbing study, they chose the farmers market because it was a busy place and easy to remain unobtrusive (Franz, 2014). I'm not watching you; I'm just looking for good deals on pickles and berries!

Reactivity: When people change their behaviors because they realize they're being watched.

Participant observation: A technique used during naturalistic observation where researchers covertly disguise themselves as people belonging in an environment.

(a) (b)

In the movie *21 Jump Street* (a), two young police officers go undercover pretending to be high school students. In *Imperium* (b), Daniel Radcliffe's character works for the FBI to infiltrate a White supremacist group. If any of them had been psychologists doing research with this undercover technique, it would have been called participant observation.

COLUMBIA PICTURES/Album/Newscom; ATOMIC FEATURES/GREEN-LIGHT INT/GRINSTONE ENT/SCULPTOR MEDIA/Album T51/Newscom

Surveys

Just ask.

Probably the most common research methodology in psychology is the **survey**, where information is collected as participants answer direct questions. Surveys can be an efficient method to collect data, especially if constructs are measured in a reliable, valid way. Surveys are often the only way to access people's private thoughts and behaviors. If the survey is convenient (for example, online) and people are compensated for their time (maybe with money or extra credits in a class), then data can be gathered from hundreds of people in just a few hours.

Survey: Asking questions directly to participants in order to collect information.

This course is full of examples of survey studies. But to stay with the social media theme, consider a study that explored whether posting "selfies" to social media signals a narcissistic personality (Barry et al., 2017). The researchers operationalized narcissism by asking 128 college students to rate themselves on statements such as "I am apt to show off when I get the chance" on a scale of 1 to 7, with 1 being *not true at all* and 7 being *very true*. This particular scale measured both "grandiose narcissism" (which involves exploiting others; the individual enhances how others see them) and "vulnerable narcissism" (which involves constantly seeking validation from others).

Participants also gave researchers access to their Instagram accounts so researchers could count and analyze their posted selfies.

Results indicated that people who scored high in grandiose narcissism (exploiting others) tended to post more selfies showing them surrounded by friends and having fun (Barry et al., 2017). People who scored high in vulnerable narcissism (wanting validation) tended to post selfies highlighting their attractive physical appearance. However, before you label yourself a narcissist for posting a selfie, their general finding was that almost everyone on Instagram posts selfies. Again, the research cycle generates new questions further research could explore: Does Instagram tend to attract narcissists?

There's a catch, or a disadvantage to surveys, and it's a big one. For the research to be valid, people must be honest. **Social desirability** influences people to respond with what they think is socially expected or what they think the researcher wants to hear. The "just ask" approach that makes surveys convenient assumes that people can and will tell the truth about themselves. But people often fib or stretch the truth when it might make them look bad.

Social desirability: The tendency for participants to provide dishonest survey answers because they want to look good to the researchers or to themselves.

On the other hand, don't underestimate the creativity of a motivated researcher! To counter social desirability, researchers sometimes embed "liar scales" in surveys. These items typically ask people to self-report common but socially disapproved behaviors—like littering or gossiping. Since many people do these things, not admitting to them may be a sign that people are stretching the truth to make themselves look good. So, would you admit to these behaviors on a scientific survey? You can try it yourself in the *What's My Score?* feature. This time, you'll have to simply choose "true" or "false" for each statement to get your score.

WHAT'S MY SCORE?

Measuring Social Desirability

Instructions: Social desirability scales are sometimes included in survey research to see if people are being honest. This scale includes several statements concerning personal attitudes and traits.

Read each item and decide whether the statement is true or false as it pertains to you.

Circle "T" for true statements and "F" for false statements.

T F 1. Before voting I thoroughly investigate the qualifications of all the candidates.

T F 2. I never hesitate to go out of my way to help someone in trouble.

T F 3. I sometimes feel resentful when I don't get my way.

T F 4. I am always careful about my manner of dress.

T F 5. My table manners at home are as good as when I eat out in a restaurant.

T F 6. I like to gossip at times.

T F 7. I can remember "playing sick" to get out of something.

T F 8. There have been occasions when I took advantage of someone.

T F 9. I'm always willing to admit it when I make a mistake.

T F 10. There have been occasions when I felt like smashing things.

T F 11. I am always courteous, even to people who are disagreeable.

T F 12. At times I have really insisted on having things my own way.

Scoring: Give yourself 1 point if you said TRUE for 1, 2, 4, 5, 9, or 11. Then, give yourself 1 point if you said FALSE for 3, 6, 7, 8, 10, or 12. Then, add your points. Higher scores indicate more attempts to manage your impression on others, or a higher tendency toward socially desirable responding on self-report scales. This means you might change your answers in psychology studies to look good to the researchers (or maybe to yourself).

Source: Crowne and Marlowe (1960).

Experiments: Research designs in which researchers compare two or more groups to see how groups differ by the end of the study.

Quasi-experiments: Research designs that compare preexisting groups to see how or if they differ in response to something in the study.

True experiments: Research designs that compare groups created by the researcher using random assignment.

Random assignment: Putting participants into experimental groups by a purely chance method (like flipping a coin).

Experiments

Don't psychologists do experiments?

Yes! In fact, for many scientists in psychology, experiments are the preferred method of research. **Experiments** are research designs in which two or more groups are compared to see how they differ by the end of the study. In psychology, there are generally two kinds of experiments—quasi and true.

Quasi-Experiments

Quasi-experiments compare preexisting groups. Quasi-experiments are necessary when researchers want to compare groups in interesting ways but can't create the groups because they have already formed themselves. If you compared boys to girls, or psychology majors to education majors, or basketball players to tennis players, you'd potentially be doing a quasi-experiment. In quasi-experiments, each group usually goes through the same experimental procedure, and the researchers are interested in whether people in the groups respond differently compared to each other.

For example, one study wanted to discover what influences children's ability to identify fake news on social media. The researchers compared how well children from two different countries—the Netherlands and Romania—could detect fake news on social media (Dumitru, 2020). Researchers showed schoolchildren fake websites from the environmental group Greenpeace supposedly fundraising to save fictitious endangered animals like the tree octopus and the jackalope. Few children from either country realized the animals were not real, which suggests a lack of critical thinking when scrolling through social media. Here, the two groups (children from each country) responded similarly to each other.

"Rare photo of the endangered jackalope!" If you saw this photo on your social media feed, would you realize it was fake?

Found Image Holdings Inc/Contributor/ Corbis Historical/via Getty Images

The problem with quasi-experiments is that even if the two groups *do* respond differently, it's not always clear exactly *why*. If children in two countries are different, for example, is it due to their culture, or the different kinds of TV shows and advertising they see, or food, or religion, or maybe how their school system is structured? There are so many variables involved, we can never be sure—which is why we need true experiments.

True Experiments

True experiments help definitively answer the "why" question by comparing two or more groups that have been made equivalent at the start of the experiment. They are usually made equivalent by **random assignment**, which first involves identifying everyone who will be in the experiment and then gives each participant an equal chance of being put in any of the experimental conditions. Flipping a coin would work; so would pulling names out of a hat. Usually, we just ask a computer to do the job.

In a true experiment, assuming there is a large sample, the people in the different groups are close to identical. *Then,* each group goes through a different experience that's controlled by the researcher (their experimental "condition"). In pharmaceutical trials, one group gets the real drug while the other group gets a *placebo,* or a treatment the participant believes is real but is not. All participants are "blind," or unaware of, their experimental condition, which means they have no idea which group they're in. Later, the researcher will compare the groups' symptoms and side effects to help determine whether the drug was effective.

A true experiment about social media investigated how people evaluate each other's profiles (Antheunis & Schouten, 2011). The researchers created fictitious profiles that differed in terms of (1) how many "friends" the fake person had, (2) how physically attractive those "friends" were, and (3) whether the posts on the profile were positive or negative. Over 500 high school students were randomly assigned to read just one of the 12 fictional profiles. The results showed that the high schoolers didn't care about how many friends someone had, but their ratings of the fictional person in the profile went up if the person had conventionally more attractive friends and positive posts on their profile.

Note that there are two ways you could set up a true experiment. The example just described is what researchers call *between-participants* designs. This means that you have multiple groups, and each group receives a different treatment. For example, half of the participants view and react to social media Profile A while the other half view and react to social media Profile B. There is another option, which comes in handy if you have a limited number of participants.

You could instead do a *within-participants* design. Here, you have a single group of people who experience multiple treatments or conditions—then you compare their reaction to each condition. You might have the same group of people first react to Profile A, then react to Profile B. It's still an experiment, just all within the same group. Here, you'd have a few other things to worry about, such as whether the viewing order would matter in forming impressions of profiles. So, you might want to randomly assign some people to see Profile A first while others see Profile B first. Now, the random assignment is about *order* of conditions for each person, instead of *which* condition each person will experience.

Either way, experiments allow for comparisons between and across conditions that no other methodology can provide. They are the only method that allows researchers to make cause-and-effect conclusions.

Experimental Groups and Variable Types

You already know that experiments compare groups or conditions. Many (but not all) true experiments compare two specific types of groups. In the **control group**, participants are essentially left alone to serve as a neutral or baseline group. In drug trials, those who get the placebo are the control group. In contrast, in the **experimental group**, the participants get some intervention or change in the environment. In drug trials, those who receive the actual drug are the experimental group.

There are also different types of variables. The **independent variable** is the influencing variable at the *beginning* of a study that, according to the hypothesis, may cause the two (or more) groups to differ from each other by the end of the study. It is what is changed and controlled in a study. The independent variable is what differentiates Group 1 from Group 2 (and Group 3, and so on). In a drug experiment, the independent variable is whether the participant receives the drug or the placebo. In the social media experiment described earlier, the independent variables were the number of friends, attractiveness of friends, and positive or negative posts on social media profiles they viewed (so, three independent variables). They are called "independent" because they cannot be influenced by the other variables in the study.

Control group: A neutral or baseline group in a study, used as a comparison to what is being tested.

Experimental group: The group (or groups) in a study that experiences an intervention or change, to see how that change affects the participants.

Independent variable: A variable that's manipulated at the beginning of the study, creating groups that will be compared to each other.

TABLE 2.2

Independent and Dependent Variables in Experiments

STUDY BASICS	INDEPENDENT VARIABLE	DEPENDENT VARIABLE
Students listen to either classical or rock music while they study, to see if music affects their memory on a test later.	Type of music (classical or rock)	Performance on the memory test
People write an essay about either death or puppies, then rate how much anger they feel.	Essay topic (death or puppies)	Level of anger
Children watch a commercial with dolls or with trucks, then are rated on how aggressively they play with clay and crayons.	Commercial topic (dolls or trucks)	Level of aggression
Sports fans see images of athletes wearing black jerseys or green jerseys and are asked to rate how well they expect each player to do that year.	Jersey color (black or green)	Expectations of players' performance

Dependent variable: The measured outcome at the end of a study, to see how the groups in an experiment had different results.

The **dependent variable** is the outcome at the *end* of the study that is measured. It's what the hypothesis suggests will be the result of being randomly assigned to a particular group. The dependent variable in the drug experiment is whether the drug worked. In the social media experiment, the dependent variable is how attractive the participants found the person in the fake profile. It is called a *dependent* variable because it is the outcome that, hypothetically, depends on the participant's group placement (which condition they were in).

Table 2.2 summarizes independent and dependent variables; you will grow more comfortable with these terms as you use and encounter them in each chapter.

Confounding variables: Other explanations for why the outcome of study happened, besides what the researcher is testing.

Finally, **confounding variables** are alternative explanations for the outcomes in an experiment. They are anything that resulted in group differences besides the independent variable. In experiments, it is important to control confounding variables because they confuse the logic of cause–effect that experiments hope to reveal. Quasi-experiments have more confounding variables than true experiments, which is why true experiments are preferred. Ideally, if all the groups in a true experiment really were as equivalent as possible at the start of a study—and the *only* real difference was the independent variable—then any differences in the dependent variable must have been due to experimental condition. At least, that's the goal.

SageVantage Practice what you learn in **Knowledge Check 2.2**

CORRELATIONAL ANALYSES

Have You Ever Wondered?
2.3 What are correlations?

>> **LO 2.3 Explain how we show relationships between variables and why correlation doesn't imply causation.**

Do you think that more time on social media is associated with worse grades in college?

Once a study has been designed and the data are gathered, it's time to analyze the results.

Many psychological studies gather all their data from just one group of people. Unlike experiments that compare people in two or more groups, correlational research collects a lot of data from every person in the study, often based on multiple constructs.

For example, one survey asked participants to provide information about their time on social media as well as several other variables related to college life (Nwosu et al., 2020). The researchers found that simply measuring time on a computer didn't have anything to do with college success—because what matters is what people are *doing* on their computer. Time spent on the computer studying or doing homework was fine—but if students were spending that time on social media to procrastinate, then more time on the computer was associated with worse academic performance.

A **correlation analysis** tests whether two constructs (or variables) you're measuring are systematically tied to each other. Note that with correlations, you don't manipulate anything—you just measure. Knowing the score on one variable makes it possible to predict a score based on the other variable. Most correlational variables are continuous, meaning they're measured on a continuum or range. Let's imagine that both your time on social media and time spent studying have a range of possible data points between 0 and 24 hours per day. Similarly, someone's cumulative GPA could be measured continuously, but this time on the traditional range from 0.0 to 4.0.

Correlation analysis: Statistical analyses testing whether two variables are systematically tied to each other.

Scatterplots

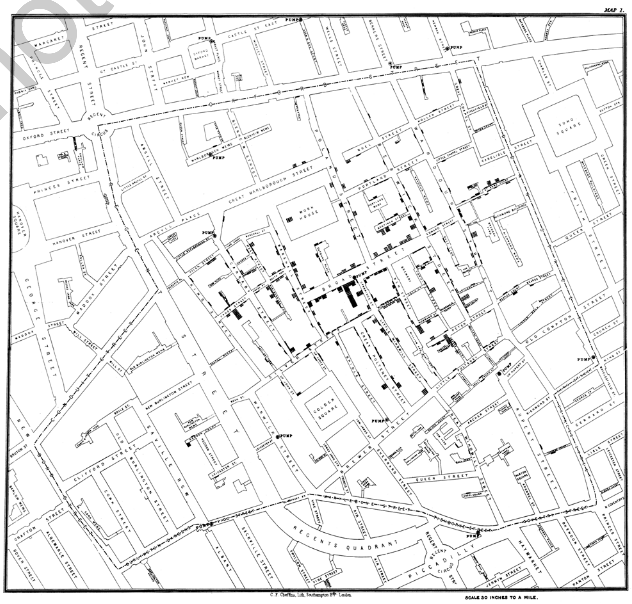
Visualizing a data story can save lives.

Before Francis Galton devised the mathematical formula for a correlation (in 1888), English physician John Snow drew a map showing how two ideas were associated with each other (Figure 2.2) that helped explain London's terrifying cholera epidemic of 1854 (Brody et al., 2000; Stigler, 1989). The map demonstrated that there were more cholera deaths among families living closer to a certain water well on Broad Street. Each dot represents a real person who died after drinking water from that specific well, which helped Londoners realize how cholera was being spread through the community. Though a bunch of dots may seem trivial, don't underestimate the power of a data story effectively told with a well-designed picture, image, or graph.

FIGURE 2.2

John Snow's Map of the Cholera Epidemic

Creation of a map showing locations affected by the cholera epidemic of 1854 in London helped people identify the cause and how the illness was spread through water.



Source: Snow (1854).

Scatterplot: A graph used to show a pattern between two continuous variables (a correlation).

Graphs used to show patterns between two continuous variables are called **scatterplots**. So, scatterplots are visual representations of a correlation. Figure 2.3 demonstrates a correlational data story framed within a horizontal x-axis and a vertical *y*-axis. The orange line summarizes the group's overall data story: As study hours increase, so does GPA.

Like John Snow's map, each dot represents two pieces of information about a real person. The line shows a general pattern and is called the "line of best fit." The overall pattern is that more studying is usually associated with better grades. The people far away from the line are called *outliers,* or cases that are substantially different from the overall pattern. Find the person in Figure 2.3 who rarely studies yet has a 3.5 GPA—and then find the student who studies about 25 hours per week to earn similar grades. A visual correlation shows both what is typical and that there are exceptions due to a variety of potentially interesting factors. Creating a scatterplot is usually the first step to correlational analysis.

Positive and Negative Correlations

There are two types of correlations.

Correlations in which both variables move in the same direction are called *positive correlations*. As one goes up, so does the other—and vice versa. For most students, more time studying predicts higher grades (measured as GPA). Positive correlations such as Figure 2.4 display a line rising from the lower left to the upper right. In a *negative correlation,* that line moves downward from upper left to lower right because the two variables move in *opposite* directions. As the hours spent procrastinating on social media increase, GPA goes down (Nwosu et al., 2020).

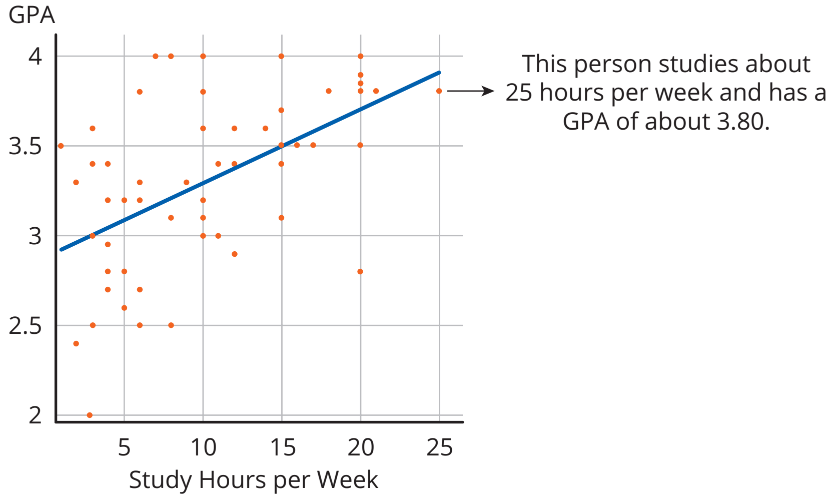
Correlations can be summarized as a number ranging from negative −1.00 to positive +1.00:

- If every data point falls exactly on the line, then there is "perfect" correlation between the two variables.

FIGURE 2.3

Sample Scatterplot

In this graph, each dot represents one person. For each person, study hours per week fall on the *x*-axis, and grade point average (GPA) falls on the *y*-axis. By looking at the general pattern, we can determine whether the two variables are correlated.



- The number summarizing that correlation will be positive if the line is rising upward from left to right and negative if moving downward from left to right.
- As the number gets closer to zero (such as –.01 or +.01), the dots look more like a cloud with no clear pattern—there is no apparent correlation between the two variables.

Figure 2.4 shows examples of how you can think about correlations at different strengths.

A Warning About Correlations and Causation

There are three things to understand when evaluating if two related variables are associated with each other.

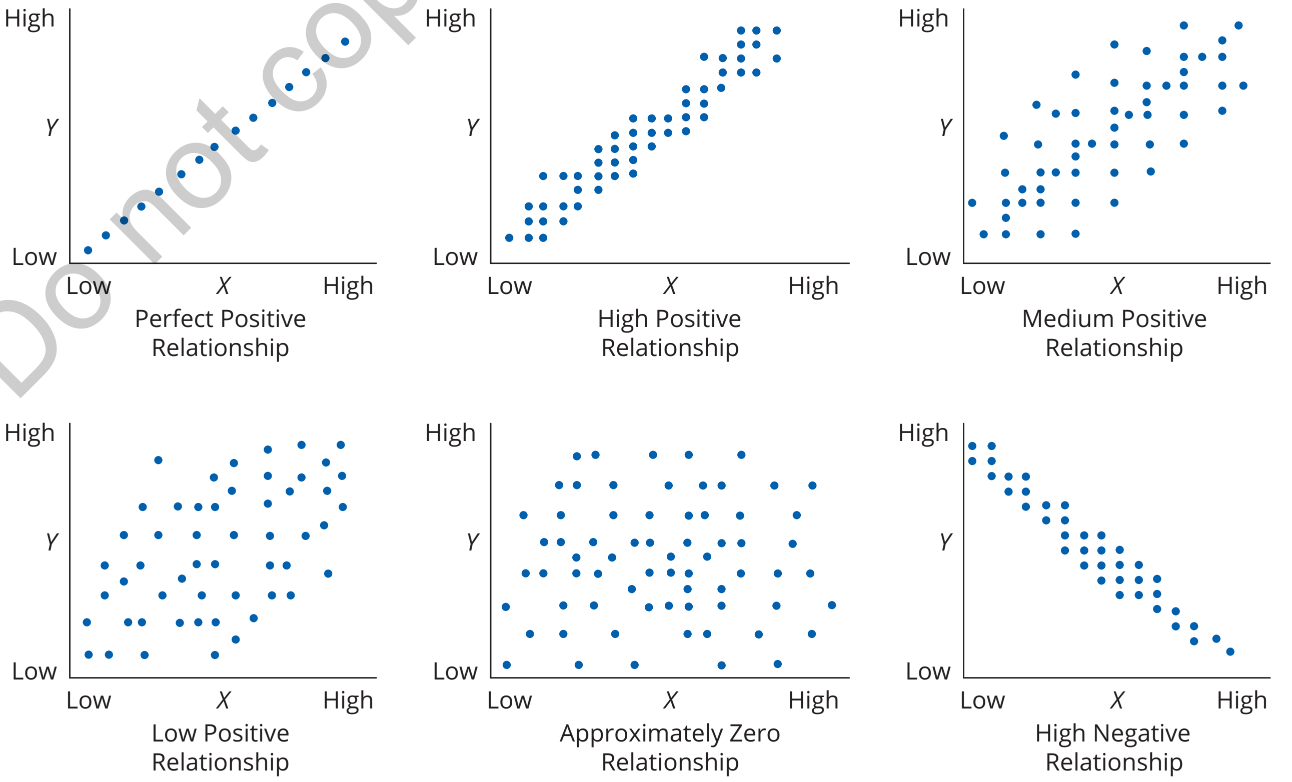
Caution 1: *Correlation does not imply causation.* If two things keep happening at the same time, it does not mean that one is causing the other. A third variable could also be the cause. In the case of a student who spends many hours studying and has a very good GPA, both outcomes might have been caused by the student's (1) motivation to do well, (2) level of pressure from parents, (3) amount of enjoyment of class subjects, or (dare we hope) (4) the skill and engagement of talented professors.

FIGURE 2.4

Correlation in Graphs

Correlations always range from –1.00 to +1.00. The sign (positive or negative) indicates whether the two variables move in the same direction or in opposite directions.

The number (from 0.0 to 1.0) tells you how well each data point fits onto a general pattern. If a correlation is zero, it means there is no pattern or association between the two variables.



Caution 2: *Cause and effect can occur in either direction.* Sure, it seems logical that if someone spends all their time on social media, that might cause their grades to go down. But it could work in the other direction. Their grades might go down first, which the student finds so discouraging that they then escape into the distracting comforts of social media. In this scenario, poor grades were the cause and more time on social media the effect.

Caution 3: *Spurious correlations are common.* Many millions of things are happening in our world at any given time. Consequently, there are unlimited combinations of spurious, or false, correlations where variables appear related but actually represent meaningless patterns. Tyler Vigen (2015) has some fun with this insight by reporting ridiculous, spurious correlations, such as the correlation between the number of Nicholas Cage films in a given year and how many people drowned in swimming pools the same year ($r = +0.66$), or letters in the final word of a Scripps spelling bee and deaths from venomous spiders that year ($r = +0.81$). It's pretty unlikely these variables are actually tied to each other in any meaningful way, despite relatively high correlations.

Remember—spurious correlations can trick you into believing that outlandish conspiracy theory reflects reality. Two things happening at the same time does not mean that one is causing the other. *You* need to be the critical thinker—sometimes correlations are life-saving clues that can help save lives from terrifying epidemics. Other times, they're just coincidences.

Because of these important cautions, some researchers prefer to set up experimental methods and analyses—the topic of our next section.

Sage Vantage — Practice what you learn in **Knowledge Check 2.3**

EXPERIMENTAL ANALYSES

>> LO 2.4 Compare and contrast *t*-tests and ANOVAs.

? Have You Ever Wondered?
2.4 How are the results of experiments analyzed?

Unlike correlations, true experiments provide solid evidence about cause and effect.

Remember that a true experiment relies on random assignment to groups because it makes each experimental condition equal before the experiment begins. The *only* possible difference in the outcome between competing groups is the independent variable.

We have Guinness to thank for the statistic known as the *t*-test.

iStock.com/WaraJenny

Imagine you are conducting an experiment that tests students' multitasking ability by listening to a lecture while browsing their favorite social media. How do you discover whether spending more time on social media during a lecture (the cause, the independent variable) influenced exam scores (the outcome, the dependent variable)?

- Each student will be tested on the content of the lecture at the end of class.
- You recruit 100 students and randomly assign 50 to look at social media for 30 minutes during the lecture, while the other 50 look for 10 minutes.

You've got the data; now you must analyze the results. But first, let's take a break for a mental beer.

Comparing Two Groups: The *t*-Test Statistic

For the rest of your life, whenever you think of beer, think also of the *t*-test.

Think of Guinness ale, to be specific. You don't have to drink alcohol or like beer to understand why the creation of the *t*-test was a such a powerful industrial secret. The ***t*-test** is statistical analysis that compares averages and ranges of two groups, to see if they are different from each other. William Sealy Gossett, a chemist and statistician who also happened to be the head brewer at the Guiness brewery in Dublin, Ireland, developed the *t*-test to monitor the quality of morning versus afternoon productions of Guinness without having to sample every keg (Mankiewicz, 2000). It would not be good for Guinness's business if the morning batch tasted different from the afternoon batch. #thanksbeer!

***t*-test:** Statistical analysis that compares the outcomes of two groups, to see if they are different from each other.

There were two critical components to Gossett's statistical invention:

1. The samples of beer each morning and afternoon had to be random (to avoid bias).
2. The number of samples had to be big enough to fairly represent all beer casks Guinness made during a production run.

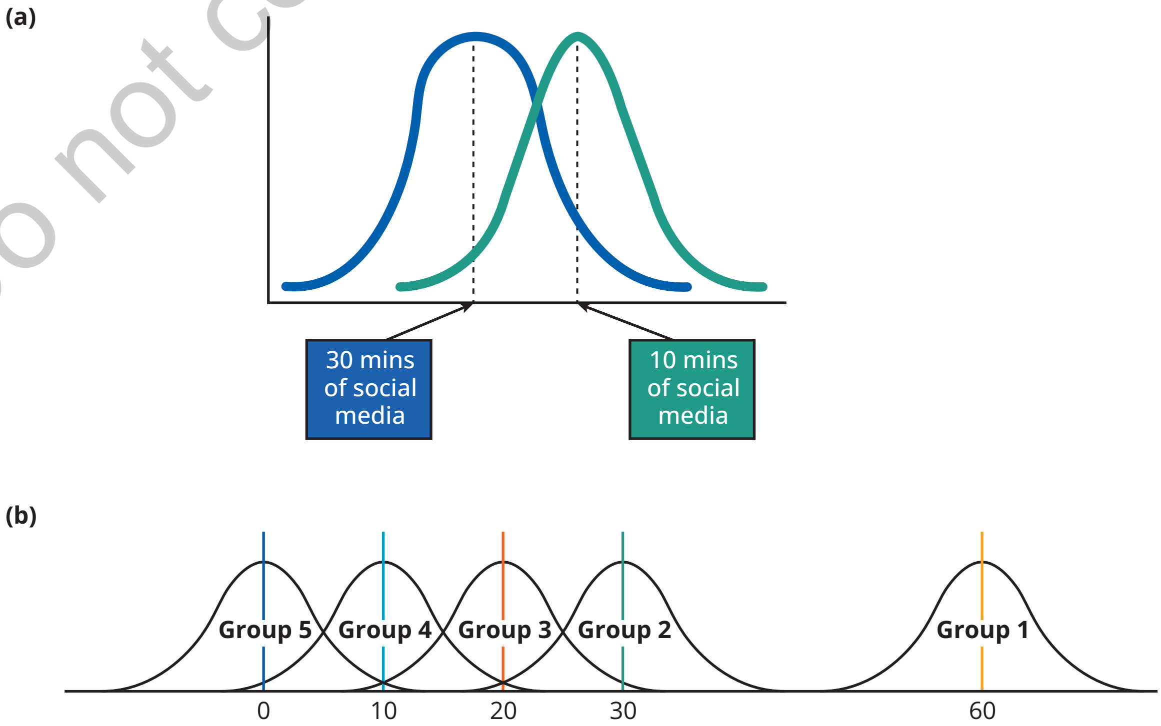
Since other manufacturers faced similar challenges to quality, Gossett was prohibited from sharing his powerful industrial secret that ensured quality control of the manufacturing process. But he published it anyway, under the anonymous penname of "Student"—and without mentioning beer. Statisticians and some textbooks still refer to it as "Student's *t*." It would be kinder to call it "Gossett's *t*."

Let's go back to your social media and multitasking experiment: Half the class pays attention to social media for 30 minutes, the other half for 10 minutes. The test scores of all participants are visualized in Figure 2.5 and organized into the two

FIGURE 2.5

t-Tests and ANOVAs to Compare Groups

One way social scientists look for patterns is by comparing average scores within different groups of participants. When we compare two groups, as here, we use a *t*-test. When we compare three or more groups, we do an analysis of variance, or "ANOVA."



groups: 30 minutes in blue, 10 minutes in green. Note that overall, the green group (10 minutes) did better on the test than the blue group (30 minutes).

Comparing Three or More Groups: Analysis of Variance

What if you have more than two groups that you want to compare?

In the example experiment, you might have set it up with five groups: one with no time at all on social media during the lecture (a control group), a group allowed 10 minutes, another group that used social media for 20 minutes, one who used social media for 30 minutes, and another that was on social media for the entire class. To analyze the results of an experiment with more than two groups, the same logic as the *t*-test applies, but the mathematics are slightly different. **Analysis of variance** (**ANOVA**) compares three or more groups and tells us if one group is significantly different than the others (see Figure 2.5). If so, we must dig further to find out which group(s) stand out.

Analysis of variance: Statistical analysis that compares the outcomes of three or more groups, to see if they are different from each other.

ANOVA: See *analysis of variance.*

Sage Vantage — Practice what you learn in **Knowledge Check 2.4**

ANALYZING THE QUALITY OF RESEARCH

>> LO 2.5 Summarize how studies can be analyzed for quality.

Have You Ever Wondered?
2.5 How can I tell if a study is done well?

How can you tell if a study is done well?

Evaluating research is like buying a car. Quality control matters, and it's important to ask educated questions. As you encounter research studies during this course, ask the same questions you would when buying a car. Does it get reliable ratings from reviewers? Are the people selling it trustworthy? Does it sound too good to be true? Are you vulnerable because you are desperate to believe in some key feature?

The history of science tells us that sometimes great ideas are ignored and terrible ideas embraced with dangerous enthusiasm. Whether it's a used car, a potential romantic partner, or a scientific study, the key to making a good evaluation is an attitude of hopeful but healthy skepticism. There are three important questions you should ask when evaluating research.

Random Sampling—Don't Be WEIRD

Who were the participants?

All published studies have a section describing who or what was in the study. Researchers start by identifying their population of interest, the group of people they want to know about. It might be people who have a certain mental disorder, people of a certain age, or students at a specific school. It's unlikely that *every* person in the population will participate. So, the "sample" refers to just the portion of the population that participates in the study.

Random sampling: Method of choosing who will participate in a study from the larger group of interest in an unbiased, random way.

The best way to avoid a biased sample is by **random sampling** from the population (*not* the same thing as random assignment to groups). Random sampling means that the people in the study were randomly chosen from the larger population of interest—everyone has an equal chance of being selected. This is often accomplished by a computer generating a random sublist of names.

Random sampling is the ideal because the sample will be a miniature version of the population. For example, if the larger population is 25% business majors, 25% education majors, 25% science majors, and 25% humanities majors, a truly random sample will produce very close to the same percentages. When the sample represents the population, then the results from the sample are **generalizable**—or probably true, in *general,* for the larger group.

This family appears to be "Western," Educated, and from the United States (an Industrialized nation). They also appear to be relatively Rich and from a Democratic country. That means the acronym "WEIRD" applies to them. If they were the only people used in psychology studies, then we'd know a lot about WEIRD people but very little about most of the other people living in the world.

iStock.com/ljubaphoto

But the real world of research often falls short of the ideal. Psychologists often resort to a convenient sample by simply asking for volunteers or offering an incentive like extra credit to students. Thousands of studies in psychology rely on people between 18 and 22 years old who are attending colleges in the United States. That may limit how generalizable the studies are. Can these results apply to younger people, older people, people without the privilege of going to college, or people in other countries?

This problem is sometimes known by the acronym WEIRD. WEIRD samples are those where the participants are:

- "Western" (from places like the United States),
- Educated, and from
- Industrialized,
- Rich,
- Democratic cultures.

Generalizable: A term describing studies in which the sample of participants represents the diversity in the larger population of interest.

We know less about other kinds of people—the diversity in participants is sometimes lacking. This is one reason cross-cultural research is so valuable to psychology and why researchers work hard to increase the diversity among participants. Limited diversity in a sample is also why a study might not replicate when tried again. This is one reason to consider the ideas in the next section—Reliability and Validity.

Reliability and Validity

Was everything in the study done correctly?

Always ask about the reliability and validity of the study. **Reliability** refers to whether the measurements were consistent. Over time or across multiple tests, would participants get the same score (on a variable assumed to be stable)? If there are strange fluctuations in how the constructs were operationalized (defined and measured), then the results might be questionable.

Reliability: Whether the measures in a study are consistent over time and place.

Several kinds of validity ensure a study stays focused on what it was intended to study in the first place. Let's consider three important kinds:

- **Internal validity** means the internal structure of the study was set up correctly; the results mean what we think they mean. Are we measuring social media use, specifically, or just time looking at the phone? Methods like random assignment to groups increases internal validity—and our

Internal validity: Confidence that a study was designed correctly and the results mean what we think they mean.

confidence in the connections between the independent and the dependent variables—because it helps rule out alternate explanations.

Construct validity: The degree to which tests, surveys, and so on chosen for a study really measure what we think they're measuring.

- **Construct validity** is whether methods are measuring the concept that we started with. A poorly written survey might confuse participants; a lie detector measuring palm sweating might only measure nervousness. Those problems would decrease confidence that we were measuring what we intended to assess.

External validity: The extent to which results of any single study could apply to other people or settings (see *generalizability*).

- **External validity** refers to whether the results apply to other people or settings. For example, the results from using a convenient sample of college students to study recognition of famous hip-hop music probably would not apply to people older than 70 who live in a culture that doesn't listen to much hip-hop. External validity can also be questionable if the study is so artificial or strange that it doesn't apply to behaviors outside of the lab. It would be strange if a researcher asked you to scroll through social media while riding an elephant, listening to a K-pop band on headphones, and counting the number of times you chew your food before swallowing. It's unlikely to have much meaning or usefulness to psychology (or any other science).

The CRAAP Test

Is this study CRAAP?

The CRAAP test is an efficient (and memorable) way to evaluate any study. The guidelines in Table 2.3 were first suggested by librarian Sarah Blakeslee (2004).

The CRAAP test can help you assess blogs on social media (Wichowski & Kohl, 2013). For example, you could consider its purpose in the context of its content—whether the blog is a personal story to help others, informal journalism, fame-enhancing, or sharing scientific advances with the public. In a world of malicious "fake news," a CRAAP test can help you evaluate whatever is popping up on your screen.

TABLE 2.3

The CRAAP Test

CRAAP	QUESTION
Currency	When was the information published? Is it current and still relevant to people today?
Relevance	Does this relate to your topic of interest? Who is the intended audience for the published paper?
Authority	What are the researchers' credentials? Is contact information available? Is the publisher reputable?
Accuracy	Are the conclusions supported by the results, or are they a bit of a stretch? Are the conclusions free of bias or emotion? Has the study been reviewed by other scholars in the field (a process called "peer review")?
Purpose	Is the purpose of the study science, or it is promoting a product to sell? Is it fact, or propaganda? Are there political, religious, or personal biases in how the researchers suggest the results should be used?

Sage Vantage — Practice what you learn in **Knowledge Check 2.5**

THE OPEN SCIENCE MOVEMENT

>> **LO 2.6 Explain how the open science movement is addressing the replication crisis in psychology.**

Have You Ever Wondered?
2.6 What is the open science movement?

Psychology has evolved.

Early psychologists did many things now considered unscientific or unethical. Unfortunately, so have some recent psychologists. If psychology is to make any progress, then the rising generation (you!) needs to be open, honest, unbiased, and ethical. A recent controversy rocked the psychology world so hard that some people quit the field entirely.

The Replication Crisis

Were the studies a bunch of lies?

In addition to reliability and validity, science relies on **replication**: getting the same findings over and over again. If a psychological principle is strong and real, then we should see the same outcomes across different settings—especially in studies conducted by *other scientists*.

Replication: Getting the same findings over and over again, with different participants, in different settings, and with different researchers.

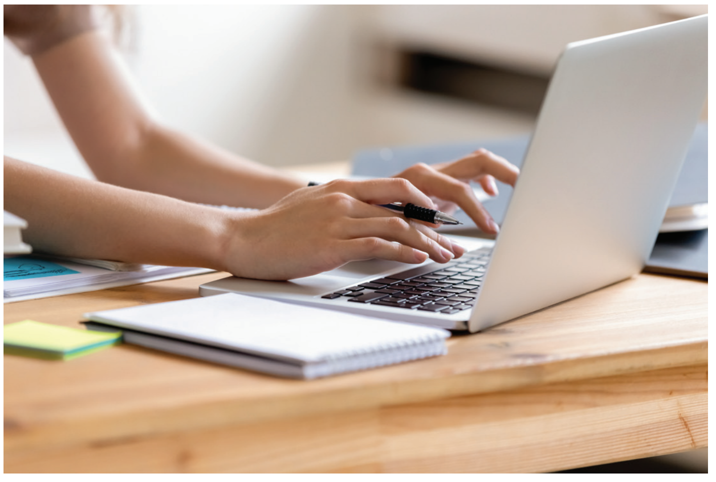
Over 250 psychologists tested the replicability of 100 classic studies (Camerer et al., 2018; Diener & Biswas-Diener, 2019; Edlund et al., in press; Open Science Collaboration, 2017; Yong, 2018).

They found a bombshell: They could replicate only 40% of the findings from the original studies. This problem was quickly labeled the **replication crisis** in psychology. Psychology is not the only discipline confronting this issue, but our troubles got the most attention. How could this have happened?

Replication crisis: The controversial finding that only 40% of several classic psychology studies replicated years later.

Some blamed elements of academic culture, such as the pressure for professors to publish their research results to keep their jobs. Others accused specific psychologists of shoddy work (at best) or outright lying (at worst). Still others identified issues such as publication bias from academic journals or private companies paying for studies that might promote their products. A few researchers quit psychology, others lost their jobs, some received death threats, and others started a crusade to weed out dishonest researchers (Cairo et al., 2020; Lilienfeld, 2017; Renkewitz & Heene, 2019; Singal, 2017).

Bullying and death threats are not appropriate responses to any situation. What might help is critical thinking about ways to make psychology better (Schooler, 2014). First, there are some reasonable explanations for low replication rates. Cultures change, so results from studies performed in 1970 might not be the same if the same studies were performed today. Another explanation is that slight differences in a study's setting may affect participants without the researchers' knowledge (e.g., Mischel, 2014; Mischel & Ebbesen, 1970; Watts et al., 2018). That doesn't necessarily mean the original study wasn't valid at the time or poorly done. But it does mean we should ask better questions. A failed replication might directly affect your life in college. To see how, check out the *Spotlight on Research Methods* feature.



Does it matter if you take notes on a laptop or in a paper notebook? See the *Spotlight on Research Methods* to learn about some research over this question.

iStock.com/fizkes

SPOTLIGHT ON RESEARCH METHODS

Understanding Replication: Should You Take Notes by Hand?

As scientists, we obsess about getting it right. This impulse is not entirely about being right in the sense of seeing hypotheses supported (though that is nice) but more about getting the facts straight—even if we don't like those facts.

Lots of college and university professors were thrilled when a study claimed that memory for classroom material was far greater when students took notes by hand—with a pencil or pen, on actual paper—than by typing (Mueller & Oppenheimer, 2014). No more annoying typing noises in class! No more temptation to check your social media or go shopping. Simply go "old school" and take notes on a piece of paper and learn more along the way.

In the study, students were showed five TED talks:

- The students took notes either on laptops or paper notebooks.
- Next, the students did distracting tasks for 30 minutes.
- Finally, they were tested on their understanding of the TED talks.

Results showed that both note-taking groups performed well on basic memory questions (no difference there). On harder conceptual and application types of questions, people who took notes on paper tested significantly better. The authors (Mueller & Oppenheimer, 2014) wanted to be confident with their results, so they did three studies—and each showed the same pattern.

But wait. A few years later, another paper questioned those results. In a direct replication of the original study's method, new researchers showed that for their participants, taking notes via typing or longhand writing just didn't really matter (Urry et al., 2021). They also pointed out eight other studies showing no real difference. They even called their study, "Don't ditch the laptop just yet."

So, do you believe the results from the original study? They failed to replicate—does that mean it doesn't really matter how you take notes? Or is there more to the story? Only additional research will be able to tell us for sure.

Regardless of where people fall on their view of the replication crisis, it was an opportunity for psychology to do better. Just in the past decade or so, the field has majorly changed how many studies are conducted and published. The goal is to make psychological research as open and honest as possible.

The Open Science Movement

How can we make psychological science even better?

Open science: A movement to make science more transparent, cooperative, reproducible, and honest.

Open science is a movement to make scientific research transparent, accessible, cooperative, reproducible, and honest. The aim is to remove any barriers in the study's creation, analysis of the data, sharing of the results, or understanding of the conclusions. Open science is a way of asking other scientists to form a team together in a transparent, honest environment. One specific goal is to increase the number of studies focused on replication of previous work, so we can be confident in the conclusions we make and in the theories we teach in classes and textbooks (like this one!).

There are three practical consequences for psychology resulting from the open science movement: preregistration, results-blind peer review, and publication badges. Find out more about this exciting trend by searching online for:

- The Center for Open Science
- The Open Science Framework
- OpenScience

- ORION Open Science
- The FOSTER Portal

Preregistration

Exploratory research occurs when a scientist doesn't have a hypothesis—they're just curious about some variables that might be related. However, scientists could publish such studies and pretend they had predicted the outcomes from the beginning. They look super smart! But it's not honest.

Open science's solution to this form of playing pretend is *preregistration,* or sharing hypotheses, procedures, and statistical plan for analyses in advance of collecting data (see Nosek et al., 2017). Several preregistration templates help researchers through this process; independent websites allow researchers to publicly declare the details of their studies beforehand.

This practical solution also has some practical problems. You might commit to having 100 participants but can only get 75, or participants can misunderstand procedures despite being presented with well-articulated plans. A typo may change the meaning of the instructions. These practical problems also have another practical solution: honesty. Document and report your mistakes and judgment calls, explain why you made them, and discuss how you addressed them.

Results-Blind Peer Review

Every research field has professional journals (like magazines), where researchers publish their results. Most journals are peer-reviewed—experts in the field evaluate each study prior to possible publication. Peer reviewers give anonymous feedback to the authors, suggest changes, and make recommendations—including whether to publish the article.

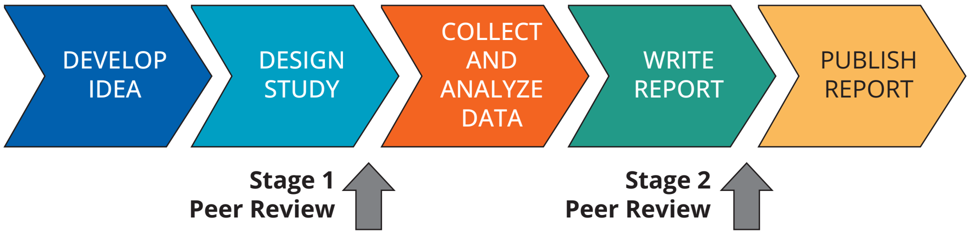
Prior to the open science movement, all this reviewing happened *after* a study was completed. Peer reviewers knew how the research story ended, which invited biases. For example, a researcher may prefer (and review more positively) studies based on a theory they favor—or vice versa. A more common, but hidden, problem was well-conducted studies that never appeared online or in print because they didn't find any differences among groups, an issue known as the *file drawer problem* (because the study went nowhere except in a file inside the researcher's office). When these unpublished studies were failed replications, it made it impossible for the field to update its knowledge.

The open science movement proposed a solution: the results-blind peer review described in Figure 2.6. Peer reviewers assess a study's procedures *before* it is

FIGURE 2.6

The Results-Blind Peer Review Process

When an article goes through the "results-blind peer review" process, outside experts give feedback about the quality and importance of an article before the data are actually collected. Then, they review a second time, focusing on whether the study followed the original design plan.



Source: Center for Open Science (2021). Retrieved from https://cos.io/rr/. Licensed under CC BY 4.0.

conducted—if they agree it has merit, then they accept it for publication. Reviewers will also provide feedback after the results are known—but now they comment on whether the study followed the preregistration plan and interpreted everything correctly. That way, even if the results surprise everyone, the study still gets published.

Publication Badges

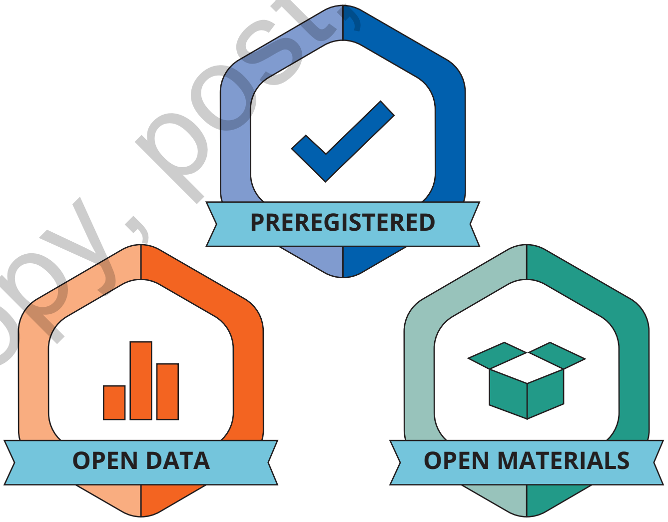
Doing good science is its own reward. There are also extrinsic incentives for participating in the open science movement. Badges are visual icons that signal researchers' participation in these practices. You can see some of the badges in Figure 2.7.

Over 50 journals now use the badge system, and early trends indicate that they do increase the number of scientists who participate in open science (Kidwell et al., 2016; Rowhani-Farid et al., 2017). More important is that the open science movement provides ethical guard rails that limit the human tendency toward self-deception and help science stay in the lane of objectivity and transparency.

FIGURE 2.7

Examples of Open Science Badges

Professional journals are increasingly marking studies with these images, called "badges," when they follow open science guidelines. These examples are from the Center for Open Science.



Source: Center for Open Science (2021). Retrieved from https://cos.io/our-services/open-science-badges/. Licensed under CC BY 4.0.

Sage Vantage — Practice what you learn in **Knowledge Check 2.6**

ETHICAL CONSIDERATIONS

>> LO 2.7 Describe ethical considerations for research with humans and animals.

Have You Ever Wondered?
2.7 How do ethics guide psychological research?

Ethical guidelines change over time.

The open science movement highlights ethical concerns that psychology has only focused on recently. And 100 years from now, future students and scientists might scratch their heads and wonder about us, "What were they thinking!?" The core of ethics in psychology is an attempt to avoid both short-term and long-term harm to the human and nonhuman animals in our studies. Let's start with what *not* to do.

Unethical Studies

Unfortunately, there are some ethically embarrassing studies in psychology's history.

You'll be learning the details of some particularly bad ones as they come up in future chapters. There's a range or continuum for just *how* unethical a study might be. For example, causing someone to feel sad for an hour is a violation, but it's not as harmful as causing someone a permanent physical or mental issue. Table 2.4 shows (in chronological order) studies that would be considered unethical by today's standards.

TABLE 2.4

Some Examples of Unethical Psychology Studies

These studies would all be considered relatively unethical by modern psychological standards.

REFERENCE	GOAL	STUDY DESCRIPTION
Landis (1924)	Studying facial expression of various emotions.	To get authentic expressions, the researcher • surprised them with a lit firecracker under their chair, • showed them pornography, and • asked them to put their hand in a bucket—without telling them the bucket contained live frogs. • To end, he handed them a butcher knife and a live rat and asked them to behead it. If they refused, he did it for them while they watched.
The Tuskegee Syphilis Study (1930s)	Studying the effects of untreated syphilis.	Over a 40-year span, African American men with syphilis were told they were getting treatment. Really, they were misled and medicine was purposely withheld.
Henle and Hubbell (1938)	Observing natural conversations between adults to analyze topics of discussion.	To ensure the people being observed didn't realize it, the researchers • eavesdropped in the bathroom, • secretly listened in on telephone conversations and even hid under beds in college students' rooms during parties.
Johnson et al. (1959)	Testing what causes stuttering in children.	Researchers selected six orphans of varying ages: • The children were told they were stutterers (even though they were not at the start of the study). • The label was reinforced multiple times for a few months, until eventually all six children started stuttering (see Silverman, 1988).
Humphreys (1970)	Studying casual sexual interactions among gay men.	• The researcher went to public parks and observed people who met there to have anonymous sex in the bathrooms. • The people knew they were being watched, but not that it was for a research study.
Piliavin and Piliavin (1972)	Testing when people will help in what appears to be an emergency.	• An experimenter collapsed on a moving subway car multiple times, varying whether they had fake blood coming out of their mouth or not. • Other experimenters on the same subway car then observed how many people tried to help (again, without knowing they were in a study).
Rekers and Lovaas (1974)	Testing whether gender-typical behaviors can be modified in children.	A 5-year-old boy expressing a desire to be a girl was trained by researchers and his mother: • They attempted to suppress any "feminine" behaviors and reinforce "masculine" behaviors with verbal praise, candy, and other rewards over about a year. • The researchers declared successful "treatment" when the child's behaviors were "normalized" (p. 181).
Middlemist et al. (1976)	Testing if invasion of personal space affected men who tried to urinate in public bathrooms.	• They found that men forced to urinate immediately next to another man (at the adjacent urinal) had trouble in terms of delay and flow. • The participants were timed without their knowledge as another researcher watched them via a hidden periscope in a nearby stall.
Kramer et al. (2014)	Studying how social media usage might affect mood.	Without users' knowledge, Facebook researchers manipulated over 600,000 people's newsfeeds: • For half, Facebook hid positive messages from friends; for the other half, negative messages from friends were hidden. • After doing this for a week, the researchers measured the users' own messages for positivity versus negativity and did find "emotional contagion."

Just as one example—and keeping with the social media theme for this chapter—consider a study created by the staff at Facebook (Kramer et al., 2014, in Table 2.4). They decided to manipulate users' newsfeeds for a week:

- Facebook staff purposely hid posts in the feeds of half the people in the study (who were unknowing participants) that contained happy, positive messages. For the other half of people, Facebook purposely hid negative messages.
- At the end of the week, the researchers coded the users' own posts in terms of emotional content.

As they expected, users who had been deprived of positive messages from friends over the week produced fewer positive posts themselves—and the same pattern emerged for negative messages. Again, users never realized they were in the study at all (and probably still have no idea that Facebook manipulated their experience).

APA Ethical Guidelines

What should you expect if you participate in a psychology study?

Notice first that several of the studies in Table 2.4 were done in the 1970s, when ethics debates were heating up across the sciences. In 1979, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research published a report, called the Belmont Report, with strict ethical guidelines that are still followed today. These guidelines will affect you very personally if you are given the opportunity to participate in research.

The specific guidelines for ethical research in psychology come from both the Belmont Report and from the American Psychological Association (APA), an organization of thousands of professionals in the field. The APA lists several rights for participants of research studies. The most important of which is that someone can stop participating in any study for any reason at any time. Some of the rights recognized by the APA include:

- Informed consent: Participants should be told what they will be asked to do and whether there are any potential dangers or risks.
- Anonymity: Personal information will be as anonymous as possible to the researchers and a participant's identity will not be published in any reports.
- Deception: Participants should be told the purpose and nature of the study. Deception is *only* allowed when knowing the truth would change how participants responded.
- Right to withdraw: Participants have the right to withdraw from the study at any time, for any reason, or to skip any uncomfortable questions.
- Debriefing: After the study, all participants can ask about the hypotheses, see the results, understand any deceptions, and receive contact information for their local ethics committee if they want to report concerns or ask questions.

PSYCHOLOGY AND OUR CURIOUS WORLD

Ethics, Nature, and Nurture: Three Identical Strangers

Album/Alamy Stock Photo

"Nature versus nurture" is a classic debate in psychology.

Twin studies is one of the interesting methods to study this question. If we can compare twins—especially identical twins separated at birth—we can see how they are similar and different. Any differences must be due to their environment, or nurture, since they share the exact same genetics, or nature. But it's rare to find twins separated at birth and even rarer to get everyone involved to consent to being in a study. What would happen if a researcher *created* this scenario without anyone's knowledge?

This incredible situation plays out in the documentary *Three Identical Strangers* (Read et al., 2018). Identical triplet infants were adopted by three different families—one blue collar, one middle class, and one affluent. None of the adoptive parents knew there were two other boys. The truth was that it was a secret research study planned by clinical psychiatrist Peter Neubauer. He did it to several sets of twins as well—ripping families apart to study how it would affect them.

When the triplets found each other by sheer coincidence at the age of 19, they had amazing similarities. They wore the same kinds of clothes and had similar personalities. They had all wrestled in junior high or high school, all smoked Marlboro cigarettes, and all preferred older women. And they also all suffered from mental illness. As they grew up, the boys started to show important differences as well, including how each responded to their unusual situation after it was discovered. When they contacted the adoption agency and Yale University, which had boxes and boxes of information about the secret study, they were barred from access to any of the information. And the researchers never even published their findings.

This strange—and highly unethical—case is captivating to watch in the documentary. No spoilers, but things did not go well for everyone. One of the boy's aunts poignantly summarized what happened like this: "When you play with humans, you do something very wrong. These three boys did not have happy endings." For more about the ethics involved in this study, read Hlavinka (2019).

Animal Research

What about research with animals?

There are also some sad stories in psychology's history that are now considered animal abuse. Modern standards are much stricter. The APA maintains a hotline to report abuses. Some of the critical guidelines for working with animals include:

- Justification: Research should have a clear purpose and provide results that will increase understanding of the species and/or benefit humans or other animals. The welfare of the animals must be monitored throughout the study.
- Care and housing: Animals should be housed in clean, working facilities and treated with care. No unnecessary stress should be imposed.
- Procedures: Experiments should minimize discomfort; if surgery is needed, animals must be given anesthesia. If animals are observed in their natural field environments, researchers should take care not to disturb the area.

CAREER CORNER

Meet Angie, Clinical Site Manager

What is your career?

I help monitor clinical trials from start to finish at a Clinical Research Organization (CRO). This includes making sure medical doctors at universities and standing research sites are well trained, follow guidelines, correctly execute trial procedures, and clean data before analyzed. I write reports on whether they're doing what they are supposed to do.

How was studying psychology important for success in this career?

Psychology helped me understand research methods and statistics and *why* clinical trials require certain procedures. For example, it helped me understand the importance of informed consent in all research. Studying psychology also assisted in understanding why certain clinical trials choose from certain demographics, certain ages, and even certain geographic regions.

How do concepts from this chapter help in this career?

I train teams weekly on the importance of informed consent. I learned what items need to be present on consent forms while studying psychology. I also train and educate medical doctors and study coordinators on the importance of following the protocol, which outlines the study design. When sites do not follow the protocol, I record it as a protocol deviation. Sometimes data need to be eliminated because the doctor didn't follow the procedure.

Is this the career you always planned to have after college?

I knew I always wanted to do research, but I didn't know the route to take. I originally started as a biology major but knew I didn't want to be in a lab or become a doctor, so I switched my major to psychology. It was a great decision for me. While I was a psychology major, I completed an internship with an epidemiologist who analyzed colorectal cancer rates based on geographical region around my home state. It was a great segue into clinical research. I now work with epidemiologists daily with study design.

If you could give current college students one sentence of advice, what would it be?

The sooner you learn how to grind and have grit, the sooner you'll reach your professional and personal goals.

Institutional Review Boards

Researchers are held accountable to protect participants.

All organizations sponsoring research with humans or animals—including colleges and universities—have committees that monitor the studies' ethics. There are generally two types of committees:

1. Institutional review boards (IRBs) review proposed studies with human participants.
2. The institutional animal care and use committee (IACUC) reviews proposed studies with nonhuman animals.

Each committee is made up of representatives from different departments at the organization. Your college or university's IRB probably includes faculty from the arts, sciences, humanities, and social sciences. In addition, there is often a lawyer and a nonresearcher from the community to give a layperson's perspective.

Psychologists planning a study must first get approval from their university's IRB or IACUC. This usually involves filling out lots of forms with detailed explanations of the exact purpose of the study, procedure, who the participants will be, whether they will be compensated for their time, projected benefits, and any possible harm that the participants might experience. The committee may ask the researcher questions, and they might even require modifications to the study if they have any concerns.

Modern psychology really cares about two basic things: doing science well and doing science ethically. If it were any other way, we (your authors) wouldn't be so proud to be in this fascinating, challenging, and rewarding field of science. ●

Practice what you learn in **Knowledge Check 2.7** SageVantage

CHAPTER REVIEW

Learning Objectives Summary

2.1 How is psychology a science?

>> LO 2.1 Explain how the scientific method is used in psychological science.

Psychology is a science because it uses the scientific method. The scientific method is a series of objective steps for empirically testing an idea. Psychologists use the scientific method to advance the field by observing patterns, generating hypotheses, testing them with research studies, and continuing the cycle. In psychology, abstract concepts called "constructs" must be operationalized, which means specifying how they will be defined and measured.

2.2 How do psychologists design studies?

>> LO 2.2 Compare and contrast research methodologies.

Five methodological options for research studies are archival studies, naturalistic observation, surveys, quasi-experiments, and true experiments. Each method has advantages and disadvantages, shown in Table 2.1.

2.3 What are correlations?

>> LO 2.3 Explain how we show relationships between variables and why correlation doesn't imply causation.

Correlations are statistical analyses testing whether two variables are systematically tied to each other. A "positive" correlation means the variables move together in the same direction, while a "negative" correlation means they move in opposite directions. Importantly, just because two things appear to be correlated doesn't mean we know if one causes the other.

2.4 How are the results of experiments analyzed?

>> LO 2.4 Compare and contrast *t*-tests and ANOVAs.

Experiments involve comparing two or more groups. If two groups are used, the results are analyzed with a *t*-test; three or more groups are analyzed with an analysis of variance (or ANOVA for short).

2.5 How can I tell if a study is done well?

>> LO 2.5 Summarize how studies can be analyzed for quality.

Quality of research can be assessed by considering several criteria, such as (1) were the participants found through random sampling, (2) is the sample generalizable to the population, and (3) does the study have good reliability and validity? The "CRAAP" test also asks about a study's currency, relevance, authority, accuracy, and purpose.

2.6 What is the open science movement?

>> LO 2.6 Explain how the open science movement is addressing the replication crisis in psychology.

Psychology suffered from the "replication crisis" when researchers repeated a large portion of classic studies and did not receive similar results. The open science movement is an attempt to improve how research is done by increasing honesty and transparency in each step of the research process.

2.7 How do ethics guide psychological research?

>> LO 2.7 Describe ethical considerations for research with humans and animals.

While some studies in psychology's history have been highly unethical, today's standards are fairly strict when it comes to research using humans and other animals. Colleges and universities have committees that consider the ethics of any study before it is done.

CRITICAL THINKING QUESTIONS

1. Now that you've learned more about scientific studies on using social media, has it affected your opinions or behaviors about your own usage? Why or why not?
2. Find a psychological study that investigated the effects of social media. Identify the following aspects of the study: (1) What was their hypothesis? (2) How did they operationalize the variables in their hypothesis? (3) What method did they use to test their hypothesis?
3. This chapter mentioned one paper that reviewed three studies showing better memory if students take notes by hand on paper, compared to if they type notes. But another paper showed that this finding failed to replicate. Think about your own preference for note-taking and explore two ideas for a study that might provide additional information. What other variables might matter? For example, do the results depend on student interest in the material? Or student personality? Or how quickly students can write or type? Generate two original hypotheses and explain how you would test them in a study.
4. Think a bit more about the studies described in Table 2.4. Rank order them from most to least ethical (knowing that they are all relatively unethical by today's standards in psychology). Then, explain why you came up with the order you did. What criteria did you use to make your judgments?

KEY TERMS