Less Fright, More Insight: Designed Specifically for Liberty University by Ray Crawford | 1st Edition | 9781524903770 | 2916 Copyright Property of Kendall Hunt Publishing

Chapter Five: Variables, Reliability, and Validity from Introduction to Research

Variables, Beliability and Validity

CHAPTER OUTLIME

Are You Equipped? An Introduction Reliability and Validity Are You Equipped Now? Chapter Summary APA Learning Goals Linkage

ARE YOU EQUIPPED?

Do you ever text while walking? We all do it! However, it is probably not a safe thing to do. Can you think of how we could use basic research elements to determine if texting has a negative impact on our ability to walk? Lopresti-Goodman, Rivera, and Dressel (2012) conducted a study to see if texting impacted speed of walking. Participants were divided into two groups. Half of the participants texted while walking and the other half of participants did not. As you might predict, those participants not texted while walking were found to walk at a slower pace than those participants not texting.

In this example, the two conditions served as the independent variable. An independent variable is a variable you think will produce a change or will influence the results. The researchers used two conditions because they believed texting would influence speed of walking. The speed of walking would be the dependent variable. A dependent variable is the variable observed and measured to see if the independent variable had an influence. In this chapter, we will provide more examples of independent and dependent variables. This will allow you to spot research design elements in everyday life.

As you go through this chapter, we also want you to keep in mind how the material relates to the APA goals for psychology majors. Specifically, this chapter will address the following goals:

· Goal 1. Knowledge Base in Psychology

You will demonstrate fundamental knowledge and comprehension of the major concepts, theoretical perspectives, historical trends, and empirical findings to discuss how psychological principles apply to behavioral problems.

Goal 2. Scientific Inquiry and Critical Thinking

You will demonstrate scientific reasoning and problem solving, including effective research methods.

From *Research Methods: Are You Equipped?* by Jennifer M. Bonds-Raacke and John David Raacke. Copyright © 2014 by Kendall Hunt Publishing Company. Reprinted by permission.

Goal 3. Ethical and Social Responsibility in a Diverse World

You will apply ethical standards to evaluate psychological science and practice and you will develop ethically and socially responsible behaviors for professional and personal settings in a landscape that involves increasing diversity.

AN INTRODUCTION

It might seem like this chapter is a vocabulary course. That's because it is! These words we've been using are specific terms with specific definitions. It is important that you learn these words and how to properly use them because you will be using these throughout this textbook, throughout your Research course, and throughout your professional career.

For example, words like reliability and validity are pretty common words, but in the context of research, they are very precise terms and concepts. I made up a new term, "research-babble". Research-babble happens when someone uses research words but does not use them accurately. Also included in this concept is those occasions when someone says something rather researchy sounding, but clearly they don't know what they are saying. The point of my rambling is this: Know the words, know the accurate and precise meanings, and use them properly. This takes time, so take the time you need to master these concepts.*

In order to discuss different types of methodologies in research, we must first start with the basics. In this chapter, we will focus on some of the elementary concepts of research. To begin, the chapter will introduce you to types of variables common in research. This will be followed by an overview of the importance of reliability in research designs. Finally, we will conclude with a discussion on validity and its role in maintaining sound research.

VARIABLES: INDEPENDENT AND DEPENDENT

A **variable** is an event or characteristic with at least two possible values. Put another way, variables can vary. For example, what would be the variable if we were to ask you, "How stressed are you about taking research methods?" In this example, the condition with an assigned or attached value is your level of stress. Furthermore, the amount of stress you indicate in your answer is the value associated with the variable. There are two variables essential to research. These two variables are the independent variable and the dependent variable.

The **independent variable** is the variable in a study manipulated by the researcher. It is being manipulated because it is the variable the researcher believes will produce a change in his or her study. The other variable of interest is the dependent variable. A **dependent variable** is the variable within a study that is observed or measured. Specifically, the dependent variable is the variable a researcher believes will change or will be influenced in the study. Usually, any change seen within the dependent variable is a result of the independent variable. In other words, any measureable change from the independent variable is measured is very important to the success of a research study. To help understand these two new terms, we will go through the example below.

Variable: An event or characteristic that has at least two possible values.

Independent Variable: The variable in a study that is being manipulated.

Dependent Variable:

The variable in a study that is observed or measured.

90 Introduction to Research: Less Fright, More Insight

^{*}Courtesy of Ray Crawford.

A marketing researcher has developed advertisements for television and Facebook to promote a new type of toothpaste. The researcher is looking to conduct a study to determine which advertisement is better at getting people interested in the toothpaste. The research question is, "Does the type of advertisement influence the level of interest in the toothpaste?" In order to test this research question, the researcher has participants watch the television advertisement or view a Facebook advertisement for the toothpaste. After viewing the advertisements on TV or Facebook, the researcher asks the participants how interested they would be in using the toothpaste in the advertisement by responding on a 7-point scale, with 1 indicating "not at all interested" and 7 indicating "very interested."

For this research study, the independent variable is the type of advertisement used (i.e., television or Facebook). Remember, the independent variable is the variable the researcher believes will produce a change. In this case, the researcher wanted to know if a difference would be seen in the types of advertisements (i.e., television or Facebook). The dependent variable is the level of interest in the toothpaste expressed by the participants. Keep in mind, the dependent variable is the measurable variable in a study. Thus, it is the variable where a change can be observed. In this study, the level of interest can be measured for each participant using the 7-point scale. This measurement can be used to see if there was a difference between participants who viewed a television or Facebook advertisement.

YOU TRY IT!

The Mozart effect is a psychological topic that has received much publicity over the past few decades. The Mozart effect is the theorized temporary increase in spatial reasoning abilities following listening to Mozart. The general public is very interested in this idea. In fact, when our oldest daughter was born, the hospital sent us home with a CD of classical music, instructing us to play the music to increase our daughter's intelligence. The scientific community has continuously studied the topic to determine if support for the effect can be found.



Recently, the Mozart effect was again studied by two researchers, Jones and Estell (2007). Part of the study was to assess whether listening to Mozart's music improves spatial reasoning skills in certain high-school populations. In order to determine this, the researchers conducted a simple study. High-school students were divided into two groups of participants. The first group listened to Mozart's music for 7.5 minutes and the second group did not. Following the 7.5 minutes of either listening to music or not, both groups completed a series of spatial problems tasks. Results showed that the group of participants who listened to Mozart's music had higher scores on the spatial problems tasks compared to the group of participants who did not listen to Mozart's music.

QUESTIONS

- **1.** What is the independent variable in this study? Specifically, what was the variable manipulated because it was hypothesized to produce a change?
- **2.** What is the dependent variable in this study? Specifically, what was the variable observed and measured in this study?
- **3.** According to Jones and Estell's (2007) results, was support for the Mozart effect found?

ANSWERS

- 1. In this study, music was the independent variable. It was manipulated by having participants either listen to Mozart or no music at all. Researchers manipulated the variable of music because they hypothesized it would influence spatial abilities.
- **2.** The dependent variable in this study was spatial reasoning skills. Remember, the dependent variable is the variable measured because it is believed to be impacted by the independent variable. In this particular study, the variable believed to be impacted by the music was a participant's spatial reasoning skills, which were tested by a series of spatial tasks.
- **3.** Yes, based on the results obtained from Jones and Estell's study (2007), the Mozart effect did occur. We mentioned earlier that the Mozart effect is the temporary increase in spatial reasoning following listening to Mozart. Upon listening to Mozart's music and completing the spatial tasks, participants who had listened to the music performed better than those who did not listen to the music. Therefore, in this experiment, the Mozart effect was supported.



Courtesy of Liz Downing.

SUBJECT VARIABLES

There is another type of variable we want to discuss that is used in research and has some of the same qualities as the independent variable. This variable is known as a subject variable. A **subject variable** is a characteristic or attribute of a participant that can impact the participant's behavior or thoughts within a study. Subject variables are often traits specific to a participant, such as gender, age, or ethnicity, and these traits can influence the dependent variable.

A subject variable is similar to an independent variable in that it is predicted to influence the dependent variable and cause a change. However, subject variables are not true independent variables. This is because a researcher can manipulate an independent variable and choose how the participant will be exposed to it. Yet, a researcher cannot

Subject Variable:

A characteristic or attribute of a participant that can impact a study.

truly manipulate a subject variable. For example, let's go back to the advertising study at the beginning of this chapter. In that study, the participants had a chance to be exposed to either a television or Facebook advertisement. On the other hand, if the researchers wanted to know the influence of participants' gender on interest in the toothpaste, the researchers could not have exposed participants to the variable of gender. Participants would have arrived to the study with the subject variable as part of their identities. Although you cannot directly manipulate a subject variable like a true independent variable, it is still possible to see if the responses on the dependent variable are influenced.



YOU TRY IT!

QUESTIONS

In the following mini-research studies, can you identify which variable is the independent variable or subject variable and which variable is the dependent variable?

- **1.** A researcher is interested in determining if different exercise programs have an influence on the amount of weight participants lose. Upon arriving to the experiment, the researcher assigns the participants to different exercise programs and measures the amount of weight lost over a 2-month period.
- **2.** You predict people from different religious backgrounds will vary in terms of their racial tolerance. You have participants complete a survey gathering information on their current religion and measure racial tolerance.
- **3.** Your university wants to know if its graduates make more money compared to high-school graduates. So, the university surveys a sample of college graduates and high-school graduates and asks participants to indicate their current salary.
- **4.** A professor wants to know if the amount of time spent studying for a test influences test performance. She has participants study for either 30 minutes or 1 hour before taking the test.

ANSWERS

- 1. The exercise programs are the independent variable. This is because the researcher manipulates the exercise program given and predicts it will influence weight loss. The dependent variable is the amount of weight lost. This is the dependent variable because it is observed and measured for change.
- 2. The variable of religion is a subject variable in this study. Religion is like an independent variable in that you can predict it will influence racial tolerance. However, you cannot directly manipulate a participant's religion. The dependent variable is racial tolerance.





- **3.** The variable of level of education is a subject variable in this study. Researchers can examine the influence of a participant's educational level (e.g., whether they graduated from high school or college) but cannot directly manipulate this variable. The dependent variable is job salary. This is the variable the university measured.
- **4.** The independent variable is the amount of time spent studying. This is the independent variable because the researcher can manipulate how the participants are exposed to study time and because study time is predicted to influence test performance. The dependent variable is the scores on the test. This is where the researcher predicts the change can be observed.

TREATMENT CONDITION

Treatment Condition:

Levels or number of groups in the independent variable.

Experimental Group:

The group that is exposed to the independent variable in a study.

Control Group:

The group that is not exposed to the independent variable.

Placebo Control Group: The treatment group which is exposed to an inert substance or object

that is similar to the independent variable but has no effect. An independent variable will always have at least two conditions. These are referred to as **treatment conditions**. Typically, the treatment conditions are the experimental and the control group. The **experimental group** is the group exposed to the independent variable. In other words, the experimental group is the group of participants given the independent variable and is the group where we would expect to see a measurable change occur. The **control group** is the group of participants not exposed to the independent variable. This group does not receive the independent variable and, therefore, we do not expect to see any measurable change in the participants.

In the study on Mozart discussed above, the experimental group was the group of participants exposed to the independent variable. This was the group that listened to Mozart's music for 7.5 minutes before completing the spatial problems task. The control group was the group of participants who did not listen to any music for 7.5 minutes before completing the spatial problems task. The control group provided a comparison so the researchers could determine if listening to Mozart influenced performance on the spatial problems task. (We will discuss treatment conditions further in Chapter 7 when we discuss levels of the independent variable.) It should be noted that although traditionally a control group is not exposed to the independent variable, there is an exception. If a researcher is comparing a new method to the current (or old) method, then the current (or old) method is serving as the control group. The new method would serve as the experimental group. An example comes from one of our classes. Jenn wanted to know how teaching statistics in her traditional manner compared to a new method where students were to develop skits to explain the material. To see if there was a difference in student learning, she compared exam scores of those students in the traditional section with those students in the new method section. In this case, the traditional manner is serving as the control group. Scores from the new method group, experimental group, are then compared with the traditional group to examine the influence on learning. So, even though we typically refer to the control group as the absence of exposure to the independent variable, in some cases the control group might involve exposure to the standard or typical independent variable.

In addition to the experimental and control groups, there is another group often used as a treatment condition in research. This group is known as the **placebo control group**. The placebo control group is similar to the control group in that participants

assigned to this condition are not exposed to the independent variable. However, the placebo control group is different from the control group because the participants are exposed to a placebo. A placebo is an inert substance or object similar to the independent variable but having no direct effect. Essentially, the placebo control group acts as an insurance policy. Sometimes researchers see measurable changes in the experimental group that are not due to the independent variable but due to the participant's belief that a change will occur. Therefore, the placebo control acts as a way for the researcher to determine how much measurable change is due to the independent variable and how much change is due to the participant's belief in a change occurring. For example, if you wanted to know if caffeine influences levels of alertness, you could have three treatment conditions. The first treatment condition, the experiment group, would receive the independent variable in the form of a caffeinated beverage. The second treatment group, the control group, would receive nothing. The third treatment group, the placebo control group, would receive a beverage believed to contain caffeine. However, in reality the beverage would not contain caffeine (such as a caffeine-free beverage). You could then measure participants' levels of alertness in all three groups. Having the placebo control group would allow you to see if alertness levels changed not due to caffeine but due to the participants' belief in the influence of caffeine on behavior.



YOU TRY IT!

There are many advertisements on television for antidepressants. Have you ever wondered how a researcher knows if these drugs will reduce symptoms of depression? Drug companies conduct research studies to determine the effectiveness of medications. Take a look at this scenario. A drug company wants to know if a new antidepressant is effective at reducing the symptoms of depression. The drug company has sampled people who are depressed and placed them into an experimental group and a control group. The participants in the experimental group will receive the new drug and the participants in the control group will receive nothing. For 8 weeks, the experimental group will receive daily doses of the new drug and the control group will receive nothing. After 8 weeks, the drug company measures the level of depression in the participants again. The results show that participants in the experimental group are less depressed compared to participants in the control group. Therefore, the drug company concludes that its drug is effective at reducing symptoms of depression.

QUESTIONS

Now that we have introduced the idea of a placebo control group, we want you to (a) describe why having a placebo control group would be of benefit to this design and (b) describe what participants in the placebo control group would do during the experiment.

ANSWERS

(a) When looking at the scenario above, you might wonder if the drug really worked. It could easily be the case that those participants in the experimental group wanted the drug to work and participants' belief alone in the drug reduced the symptoms of depression. Adding a placebo control group to the design ensures observed changes in the symptoms of depression were due to the drug itself (independent variable) and not participants' belief in the drug's effectiveness. (b) Participants in a placebo control group would have gone through the same events as participants in the experimental group, taking a pill every day for 8 weeks. The pill given to the placebo control group would look exactly the same as the pill given to the experimental group, but would be an inert substance (e.g., a sugar pill). In other words, the participants in the placebo control group would have received something that looked like the independent variable but had no direct therapeutic effect. If, after 8 weeks, the participants in the placebo control group showed the same reduced levels of depression as participants in the experimental group, the researchers should question if the drug alone is causing the change. However, if the participants in the placebo control group demonstrated similar levels of depression to participants in the control group after 8 weeks, the researcher would have reason to believe the drug is effective in reducing symptoms of depression. We will return to the topic of placebos and placebo control groups in Chapter 9.

SECTION SUMMARY

- A *variable* is an event or characteristic with at least two attached values.
- When conducting a study, there are two important variables to be considered:
 - The *independent variable* is the variable in a study manipulated by the researcher.
 - The *dependent variable* is the variable within a study observed or measured.
 - A *subject variable* is a characteristic or attribute of a participant that can impact the participant's behavior or thought within a study.
 - Subject variables are often traits specific to a participant, such as sex, age, or ethnicity.
- *Treatment conditions* refer to levels or number of groups in the independent variable.
 - The *experimental group* is the group exposed to the independent variable.
 - The *control group* is the group not exposed to the independent variable.
 - The *placebo control group* is exposed to an inert substance or object similar to the independent variable but having no effect.

RELIABILITY AND VALIDITY

In the first part of this chapter, we introduced you to the concept of variables and discussed independent and dependent variables. When you read about independent and dependent variables, it is important to consider two measurement concepts. The first concept is reliability. Reliability deals with the consistency when measuring variables in a research study. Consistency is the important feature here. For example, the speedometer on Grampa Ray's motorcycle is inaccurate as far as actual speed is concerned, but it is always off by the same amount. So, since it is consistent (even not perfectly accurate) it is still a *reliable* measure of speed.* The second concept is validity. Validity is concerned with the accuracy of the measurements used to assess different variables. Consequently, Grampa Ray's motorcycle's speedometer is not *valid*, because it does not accurately measure what it is supposed to measure, speed. It is still reliable, because it is consistent even though it is consistently inaccurate. But it is not valid, because it does not give an accurate measurement of speed.* In the remainder of this chapter, we will present you with specific techniques used to establish the reliability and validity of variables. We will begin with reliability.

RELIABILITY

Reliability of a variable is very important when conducting research. **Reliability** is the consistency of your measure to produce similar results on different occasions. Therefore, reliability is primarily concerned with being able to replicate or reproduce the findings. To make sure a measure is consistent in its ability to evaluate a variable, there are several types of reliability assessments. Let's look at some types of reliability that will be useful for evaluating the reliability of an assessment tool, for example something like the multiple choice tests you have for your courses, or other assessment tools.*

The most common type of reliability assessment is known as **test-retest reliability**. When using a test-retest assessment, you give your measure to a sample of participants individually (*test*) and then again at a later date (*retest*). Usually, the testing is done a few weeks apart. In order to make sure the measure is reliable, you look at the two scores for each participant. If the measure is reliable, the individual will have comparable scores on the two points in time. For example, if we were to give an intelligence test to your class at the beginning of the semester and then again midway through the semester, we would expect to see similar scores for each person. We would not expect the scores to be exactly the same, just close. To determine whether the measure is reliable, you compute a correlation coefficient for the scores. A correlation coefficient of 0.80 in a measure is generally seen as very reliable. At this point, you do not need to know how to compute a correlation coefficient. Rather, the goal

is to have you prepared to know what to look for when reading about this procedure in research articles. One disadvantage of the test-retest procedure is that it is time consuming.

Another method which does not require the same amount of time to assess reliability is known as the **split-half method**. The splithalf method occurs when you administer a measure to a sample of participants. Unlike a test–retest method, where you would wait several weeks before giving the measure again, the split-half method uses only the results from the first collection of data.

Reliability:

The consistency of your measure to produce similar results on different occasions.

Test-Retest Reliability:

A reliability assessment where your measure is tested on two different occasions for consistency.

Split-Half Reliability:

A reliability assessment in which a measure is split into half and the two halves are compared. If the correlation is high, the measure is said to have high reliability.



^{*}Courtesy of Ray Crawford.

Specifically, you would split the measure in half. It might sound strange to split a measure into half. However, consider assessing the reliability of a 50-question multiple-choice exam. This could be done by randomly assigning questions from the exam to two groups, dividing the exam between odd and even questions, or by dividing the exam at the midpoint. You then compute scores for each half finding a correlation coefficient between the two halves. One concern is the method in which you split the halves can impact the correlation. Specifically, if you unknowingly have several similar questions in the same half and none in the other half, the correlation will be low. Therefore, many researchers use a modified version of the split-half method known as internal consistency.

Internal Consistency:

A reliability assessment similar to the split-half method. However, the splitting occurs more than once and an average of the correlations is taken.

Parallel Forms:

A reliability assessment in which a measure is divided into half and given to two groups of people. The reliability is high if each measure given is highly correlated.

Interrater Reliability:

A reliability assessment used when a research design calls for observations of an event. Two or more observers compare results from their observations. The higher the observer consensus, the higher the reliability. The **internal consistency** method is the same as the split-half method with one exception. In the internal consistency method, you repeat the split-half procedure multiple times, thus collecting multiple correlation coefficients. Then, you average the multiple correlation coefficients. This method counteracts the impact of having too many similar questions in only one-half of the split. A commonly used statistic for computing internal consistency is Cronbach's alpha. You will learn more about statistical tests in the statistics course you complete as part of your degree.*

Another way to assess the reliability of a measure is to use the **parallel-forms method**. In this method of reliability assessment, you divide a measure into two parts. Questions from the original measure are randomly divided among the two parts. You then administer both of the parts to a sample of participants. Following the administration, you compute the correlation coefficient for the two parts. The higher the correlation between the parts, the more reliable the measure is said to be. The benefit of this approach is the brief amount of time it takes to assess reliability. However, one disadvantage is you must generate a large number of questions in order to divide into two parts.

The last assessment of reliability we want to discuss is interrater or interobserver reliability. The previous assessments of reliability were focused on how you would construct a measure (such as constructing an exam in a class). However, interrater or interobserver reliability is focused on using a measure consistently in research. Interrater reliability is used when a research design calls for observations of an event. This type of reliability assessment is used to ensure the observations being made are consistent and not biased. For example, if you want to evaluate bullying behavior in children, you could observe 100 instances of bullying and categorize the bullying into one of the four categories. If you were the only researcher making the observations, there is no way to know if you consistently categorized the 100 observations. However, if there is more than one observer, you can become more confident in the ratings. Interobserver reliability uses more than one observer and the observations of the observers are compared to assess the level of agreement. In the bullying example, two or more observers would compare the categorizations of the 100 instances of bullying. If the observers were to agree 92 times out of a 100, the reliability would be 92%, which is quite reliable. However, if the observers only agreed 48 times out of a 100, the reliability would be 48% and the measure is not very reliable. Essentially, you are calculating a correlation coefficient between the degree of similarities in the observations of the observers. To be confident in the reliability of your results, you want to obtain a high degree of similarity.

The most important consideration when using multiple observers, or raters, is that the observers themselves are consistent with each other. The raters will need appropriate

^{*}Courtesy of Ray Crawford.

training to know how to observe, what to look for, how to rate what is observed, etc. So to have good inter-rater reliability when you have more than one observer or rater, these observers need to have consistency in how they observe or rate. This takes training, which also adds time and cost to your study.*

VALIDITY

As we have mentioned, selecting your variables is an important part to any research design, as is measuring them. In this section, we will talk about the concept of validity. **Validity** is defined as the ability of your measurement to accurately measure what it is supposed to measure. This is different from reliability, which is about being able to replicate scores on future instances. Let's begin with an example.

Recently, we had a student decide to pursue his graduate degree in psychology. In order to apply to graduate school, he had to take the Graduate Record Examination (GRE). In this example, the GRE score is the measurable variable. The GRE measures a person's verbal and quantitative skills. Our student took the test and did not do very well. Believing the score was an anomaly, he took the GRE a second time. Unfortunately, the student received a similar score. It was after this second exam we discovered that the student had been taking the wrong test. Specifically, the student had been studying for the GRE-General but had taken the GRE-Psychology (or GRE subject test). The GRE-Psychology exam tests a student's knowledge of the field of psychology. You can use this example to think about validity. The GRE-Psychology is not a valid measure of a person's verbal and quantitative skills. This is because the GRE-Psychology is designed to measure knowledge in the area of psychology. Thus, it makes sense why the student was complaining that the information he studied was not on the test. You can also use this example to compare the concepts of reliability and validity. The GRE-Psychology is a reliable measure. This is because the student received similar scores both times he took the test. (By the way, if you are thinking about graduate school, you will need to take the GRE. You can learn more about the GRE at http://www.ets.org/gre.

INTERNAL, EXTERNAL, AND CONSTRUCT VALIDITY

When measuring a variable, there are several types of validity to consider. The three most common types of validity are internal validity, external validity, and construct validity. Each of these types of validity deals with a different aspect of measurement. However, the commonality in all is that they are concerned with the accuracy of the measures used in a research design. The first type of validity we will discuss is internal validity. Internal validity is an important factor for independent variables. **Internal validity** is confidence in saying the observed change in the dependent variable is due to the independent variable and not due to any outside influences. This allows you to make a causal inference regarding the influence of the independent variable on the dependent variable. This is important when conducting a study because you want to be able to show the manipulation had an impact.

Rothbaum, Anderson, Hodges, Price, and Smith (2002) examined different types of therapies to relieve fear of flying. Specifically, over a 6-week period, participants were placed into one of the three groups, where each group was exposed to a different type of therapy. After 6 weeks, the researchers measured participants' fear of flying. The interesting part to this study was that the researchers again assessed participants' fear of flying one year after the conclusion of the study. Results showed the levels of fear had

Validity: The accuracy of a measure to evaluate what it is supposed to measure.

Internal Validity: It is confidence in saying the observed change in the dependent variable is due to the independent variable and not due to any outside influences.

^{*}Courtesy of Ray Crawford.



remained relatively stable since the end of the study. Therefore, the researchers concluded that the introduction of the independent variable (types of therapy) had caused a measurable change in the dependent variable (fear of flying). We mention this study to bring up the idea that researchers need to remain vigilant as to factors that might threaten the internal validity of their study. For example, do you think the follow-up results (i.e., results after 12 months) would have been different had the events on September 11, 2001, occurred during that time? The answer is probably yes. Had the results been different and the fear level was higher, the researchers would not have been able to conclude that

the independent variable had a lasting influence. There are many threats to internal validity that a researcher must be aware of when conducting research. We will discuss many of these threats later in Chapter 9.

External Validity:

The extent to which the obtained results in a study can be generalized to other settings.

The next type of validity we want to discuss is external validity. **External validity**, also known as ecological validity, is the extent to which the obtained results in a study can be generalized to other settings. When considering external validity, you examine if any changes in the dependent variable can be applied to similar events. Specifically, can the results you obtained in the laboratory occur in a real-world setting? Researchers are often confronted with problems due to external validity. This is because a large percentage of research is conducted in a laboratory environment, where the researcher can isolate a single independent variable to determine its influence on a dependent variable. However, since research is done in such controlled environments, it is sometimes difficult to know if the causal inference drawn in the laboratory will apply to the real world.

One way to combat threats to external validity is to design experiments as close to the real world as possible. For example, researchers in the area of cognitive psychology have done much research using microworld simulations and virtual computer games. Researchers continue to think outside of the box and design experiments that are as close to the real world as possible. This allows researchers to extend results obtained in the lab to the real world.

The final type of validity that we will discuss is construct validity which is perhaps the most difficult to understand. **Construct validity** refers to the likelihood that the device or scale used to measure a variable actually is related to the topic or theory of interest. In other words, does the way we measure a variable accurately capture the theoretical construct behind that variable? If we are interested in measuring college student obesity, we could ask the following questions:

Do you over	eat?							
Not at all	1	2	3	4	5	6	7	All the time
How often de	o you e	eat fast	food?					
Not at all	1	2	3	4	5	6	7	All the time
Do you eat v	egetab	les?						
Not at all	1	2	3	4	5	6	7	All the time

Construct Validity: The likelihood that the

device or scale used to measure a variable actually is related to the topic or theory of interest.

On the surface, these questions might appear as though they will measure the variable of obesity. (This is known as face validity, where the device or scale has the superficial look to assess a variable's theoretical construct.) However, what is the likelihood that participants in your study would be truthful in their answers? In addition, just because you eat a lot of food or fast food does not necessarily mean that you are obese. Conversely, eating lots of vegetables does not mean you are skinny. If these were the questions you used to assess college student obesity, you might have low construct validity.

In order to have higher construct validity, you can assess two different components related to construct validity. The first component is **convergent validity**. The logic behind convergent validity is that your measure should converge or be similar to other measures of the same variable. Therefore, to have high construct validity your measure for a variable should show similar results to other valid measures of the same variable. Going back to the example on obesity, your measure should yield results similar to that of a valid measure of obesity. The second component related to construct validity is **divergent** or **discriminant validity**. This is the opposite of convergent validity. Whereas convergent validity argues your measure of a variable should be similar to other valid measures of the same variable, divergent validity argues your measure should be dissimilar to measures of different variables. Looking at the example of obesity again, your results from your measure should look similar to other obesity measures, but should not look similar to say a measure on diabetes. By using these two related components when developing measures of variables, researchers are able to increase construct validity in their study.

SECTION SUMMARY

- *Reliability* is the consistency of your measure to produce similar results on different occasions. There are several ways to assess reliability:
 - *Test-retest reliability* is a reliability assessment where your measure is tested on two different occasions for consistency.
 - *Split-half reliability* is a reliability assessment in which a measure is split in half and the two halves are compared. If the correlation is high, the measure is said to have high reliability.
 - *Internal consistency* is a reliability assessment similar to a split-half method. However, the splitting occurs more than once and an average of the correlations is taken.
 - The *parallel-forms* method is a reliability assessment in which a measure is divided in half and given to two groups of people. The reliability is high if each measure given is highly correlated.
 - *Interrater reliability* is used when a research design calls for observations of an event. Two or more observers compare results from their observations. The higher the observer consensus, the higher the reliability.
- *Validity* is the accuracy of a measure to evaluate what it is supposed to measure. The three most common types of validity are internal, external, and construct validity.
 - *Internal validity* provides confidence in saying that the observed change in the dependent variable is due to the independent variable and not due to any outside influences.

Convergent Validity:

A type of construct validity that states that your measure should converge or be similar to other measures of the same variable.

Divergent Validity:

A type of construct validity that argues that your measure should be dissimilar to measures of different variables.

- *External validity* is the extent to which the obtained results in a study can be generalized to other settings.
- *Construct validity* refers to the likelihood that the device or scale used to measure a variable actually is related to the topic or theory of interest. Construct validity is composed of two components, convergent validity and divergent validity.
 - *Convergent validity* states your measure should converge (or be similar) to other measures of the same variable.
 - *Divergent validity* argues your measure should be dissimilar to measures of different variables.

ARE YOU EOUIPPED NOW?



Let's revisit the topic of texting from the beginning of the chapter. This time we will pose the question, do you ever text and drive? Again, we have probably all done this before. However, is this safe to do? Research suggests it is not. For the purposes of this exercise, we are going to give you a modified version of an experiment conducted by Owens, McLaughlin, and Sudweeks (2011).

Owens et al. wanted to manipulate texting conditions while driving (i.e., driving with no texting, driving while texting on a personal phone, and driving while texting using an in-vehicle texting system) to see if this influenced visual and steering behaviors of drivers. Participants completed these conditions by texting the researcher on a closed course. Results indicated that driving with no texting produced the best results, followed by the in-vehicle system and lastly the personal phone. Thus, texting does reduce performance and is a mental distraction for drivers.

For this experiment, answer the following questions:

- 1. What is the independent variable?
- 2. What is the dependent variable?
- **3.** How did the researchers try to increase the ecological validity of the experiment?

ANSWERS

1. The variable the researchers manipulated was the driving condition. Specifically, the researchers verified if the participants were driving without texting or driving while texting on a personal phone or through the use of an in-vehicle testing system. The researchers manipulated this variable because they predicted it would influence driving abilities.

- 2. The variable the researchers measured was the performance of the drivers on visual and steering behaviors. Thus, performance was the dependent variable. The researchers did find the dependent variable was influenced by the independent variable, with texting negatively impacting driving behaviors.
- **3.** The researchers increased the ecological validity of the experiment by having participants actually drive in a car. This increases the likelihood that the results can be generalized to the real world.

CHAPTER SUMMARY

- A *variable* is an event or characteristic with at least two possible values.
- When conducting a study, there are two important variables to be considered:
 - The *independent variable* is the variable in a study manipulated by the researcher.
 - The *dependent variable* is the variable within a study observed or measured.
- A *subject variable* is a characteristic or attribute of a participant that can impact the participant's behavior or thought within a study.
 - Subject variables are often traits specific to a participant, such as sex, age, or ethnicity.
- *Treatment conditions* refer to the levels or the number of groups in the independent variable.
 - The *experimental group* is the group exposed to the independent variable.
 - The *control group* is the group not exposed to the independent variable.
 - The *placebo control group* is exposed to an inert substance or object similar to the independent variable but having no effect.
- *Reliability* is the consistency of your measure to produce similar results on different occasions. There are several ways to assess reliability:
 - *Test-retest reliability* is a reliability assessment where your measure is tested on two different occasions for consistency.
 - *Split-half reliability* is a reliability assessment in which a measure is split in half and two halves are compared. If the correlation is high, the measure is said to have high reliability.
 - *Internal consistency* is a reliability assessment similar to a split-half method. However, the splitting occurs more than once and an average of the correlations is taken.
 - The *parallel-forms* method is a reliability assessment in which a measure is divided into half and given to two groups of people. The reliability is high if each measure given is highly correlated.
 - *Interrater reliability* is used when a research design calls for observations of an event. Two or more observers compare results from their observations. The higher the observer consensus, the higher the reliability.

- *Validity* is the accuracy of a measure to evaluate what it is supposed to measure. The three most common types of validity are internal, external, and construct validity.
 - *Internal validity* is confidence in saying the observed change in the dependent variable is due to the independent variable and not due to any outside influences.
 - *External validity* is the extent to which the obtained results in a study can be generalized to other settings.
 - *Construct validity* refers to the likelihood that the device or scale used to measure a variable actually is related to the topic or theory of interest. Construct validity is composed of two components, convergent validity and divergent validity.
 - *Convergent validity* states your measure should converge (or be similar) to other measures of the same variable.
 - *Divergent validity* argues your measure should be dissimilar to measures of different variables.

APA LEARNING GOALS LINKAGE

Goal 1. Knowledge Base in Psychology

You will demonstrate fundamental knowledge and comprehension of the major concepts, theoretical perspectives, historical trends, and empirical findings to discuss how psychological principles apply to behavioral problems.

Sections Covered: Are You Equipped? An Introduction, Are You Equipped Now?

Explanation of the Goal: The field of psychology uses specific concepts to account for psychological phenomena. This chapter introduced concepts such as independent, dependent, and subject variables. The exercises at the beginning and the end of the chapter helped you practice identifying these variables.

Goal 2. Scientific Inquiry and Critical Thinking

You will demonstrate scientific reasoning and problem solving, including effective research methods.

Sections Covered: An Introduction, Reliability and Validity, Are You Equipped Now?

Explanation of the Goal: This chapter covered two areas in basic research methods. First, we introduced different research methods used by psychologists. The methods included the use of an experimental group, a control group, and a placebo control group. Next, we introduced two important terms in research: validity and reliability. After learning about these terms, you should be able to (a) evaluate the validity of conclusions presented in research reports, (b) select and apply appropriate methods to maximize internal and external validity and reduce plausibility of alternative explanations, and (c) use reliable and valid measures of variables of interest.

Goal 3. Ethical and Social Responsibility in a Diverse World

You will apply ethical standards to evaluate psychological science and practice and you will develop ethically and socially responsible behaviors for professional and personal settings in a landscape that involves increasing diversity.

Sections Covered: Are You Equipped?, An Introduction, Reliability and Validity, Are You Equipped Now?

Sections Covered: An Introduction

Explanation of the Goal: Psychological explanations are complex in nature. Thus, to answer research questions, more sophisticated designs (e.g., the use of a placebo control group) are sometimes needed. In addition, to ensure that explanations are correct, reliability and validity are key.

Excerpts from APA Guidelines for the Undergraduate Psychology Major, Version 2.0, August 2013 by the American Psychological Association. Copyright © 2013 by the American Psychological Association.

