

## Variables and Measures

THIS CHAPTER DISCUSSES HOW TO DEVELOP your variables, how to conceptualize and operationalize your variables, and how to measure your variables. We will first discuss how to operationalize concepts so that they can be measured for evidence. Measurements will be discussed, including how to measure variables using four levels of measure. This chapter will also introduce two new terms: reliability and validity.

### Variables in Research Design

Variables are fluid and sometimes difficult to define and measure. For example, researchers may come up with many different ways of defining verbal abuse, such as yelling, cursing, and belittling. How you choose to specifically define particular variables is called conceptualizing variables, and how you measure a concept (variable) is called operationalizing variables. How you define and measure your variables depends on what it is you want to know or, more specifically, what your research problem statement, question, or hypothesis is.

Conceptualizing a variable refers to how we translate an idea or abstract theory into a variable that can be used to test a hypothesis or make sense of observations. Therefore, conceptualizing a variable is another way of saying how experts are defining a concept. For example, most literature defines child maltreatment as abuse or neglect. Then these broad categories are further defined into more descriptive categories. The category of abuse can be divided into physical abuse, sexual abuse, and emotional abuse. The category of neglect can be further divided into medical neglect, physical neglect, and emotional neglect. By using more descriptive categories, we

are able to capture more specific information about the type of abuse an individual has encountered.

In research, it boils down to how we measure our variables or choose to operationalize them. Operationalizing a variable refers to how we define a concept so that it can be measured. How you operationalize a variable may depend on what information you want to collect and how you collect it. Do you only want to know if an individual has experienced childhood maltreatment, or do you want more precise information such as what kinds of maltreatment? We will come back to this later in this chapter when we discuss measurements.

### Viewing and Using Variables

It is important to keep in mind that variables are fluid and that the meaning attached to a variable can differ from person to person. In other words, they evolve in how people view them and use them. The following two examples illustrate cases in which the variables of sex and race may be viewed differently by different people.

In some studies, sex is identified as either female or male, and gender identified as either masculine or feminine. Sex is usually associated with a person's biology (chromosomes and hormones), and gender is a culturally constructed concept of what is female and what is male. Jack was born in 1957. Jack married his wife, Elaine, at age nineteen and had a son at the age of twenty. Now, at the age of fifty, Jack has officially changed his name to Jane. Jane has had a sex change operation, wears dresses and a wig, and receives hormone shots. Jane is still married to Elaine and now considers this a lesbian relationship. According to Jane's official birth certificate and marriage license, Jane is still a male, while physically and socially Jane is a female. How would you as a researcher record Jane's sex? Would this be different from how you would record her gender?

Race and ethnicity can also create similar concerns. Race is commonly used as a measure for biologically based human characteristics known as phenotypes. Categories used for race are sometimes based on skin color (white, black) or other characteristics that have been labeled Hispanic and Asian, whereas ethnicity is a culturally constructed concept connected with the history, culture, and national origin that form group identities such as Irish, Jewish, or African American. What if someone has black parents but is raised in a white family? What if you can trace your family back to Ireland but do not know anything about that culture? A pressing issue is how to claim race when your father is black and your mother is Asian, as

is the case for the famous golfer Tiger Woods. A person who is biracial is two non-specified races, but what if your father is biracial (black/white) and your mother is Native American? Does that mean you are tri-racial?

### Types of Variables

Variables can be placed into three general groups: independent variables, dependent variables, and control variables.

As stated in chapter 3, the independent variable (identified as IV) is often thought of as the variable that is controlled or manipulated by the researcher. At other times, it is described as the variable that may have an impact on a change in the dependent variable. The dependent variable (identified as DV) is the variable that is changed by another variable, or is said to depend on the independent variables. For example, if we know how many alcoholic drinks a person consumes in an hour (the independent variable), we can predict his or her blood alcohol level (the dependent variable). In social work practice, the dependent variable is the variable that is being measured to determine if change has occurred. For instance, look at the question "Did the smoking cessation classes help people quit smoking?" For this research question, the smoking cessation classes (independent variable) are the intervention being introduced, and the participants' smoking behavior (dependent variable) is what is being measured.

A predictive variable is a type of independent variable, and a criterion variable is a type of dependent variable. Prediction is a special kind of relationship where one thing precedes the other and we use information about the first to forecast or predict the second. In predictive studies, variables are sequenced or are arranged in a distinct time line. In predictive studies the predictor is the first variable (the one we have information about and are using) and the criterion is the second variable (the one being predicted). Also, you could have more than one predictor and/or criterion variable. This depends on your research question or problem statement. Let's look at some predictors of child maltreatment:

- Age (IV) is a predictor of abuse (DV): younger boys are abused more than older boys and older girls are abused more than younger girls.
- Income (IV) is a predictor of child maltreatment (DV).
- Stress (IV) is a predictor of child maltreatment (DV).
- Social isolation (IV) is a predictor of child maltreatment (DV).

A **control variable** is a variable that researchers control for in a research study. When researchers **control for** a variable, this means that they subtract the effect of that variable on the dependant variable by holding the

variable constant. Researchers often use control variables as a type of theoretical insurance. That is, they may not think that the control variable will influence the dependent variables, but they include the variable just to be safe. Many control variables are demographics. **Demographics** are the physical characteristics of a population, such as age, sex, marital status, family size, education, geographic location, and occupation. For example, assume for a moment that based on the literature, you believe a person's sex affects his or her self-esteem. Your support group for parents without partners has both male and female members. Therefore, you want to control for sex when you conduct your study of the support group by including sex in your study. In other words, you want to ensure that the support group helps consumers regardless of their sex.

By including sex, you can use statistical methods to look at the response differences between males and females, or to subtract the influence of sex from the data (depending on what type of analysis you use). Another way to control for sex in your support group study would be to assign males to one group and females to another group. You could then compare their scores.

Going back to the child maltreatment literature, you can see from the following list that marital status (IV) and race (IV) were not found to be predictors of child maltreatment (DV) when income was controlled for. Similarly, the parent's sex (IV) was not found to be a predictor of abuse (DV) when time spent with the child was controlled for.

- Single parents seem to be more abusive (DV); however, the effect disappears when income is controlled for.
- African Americans are believed to abuse (DV) more often than whites; however, this effect disappears when income is controlled for.
- Mothers abuse (DV) children more often (60%) than fathers; however, this effect disappears when time spent with children is controlled for.

As you can see, the effect of a predictor can sometimes be explained by another variable. For instance, both race and marital status no longer appear to be predictors for abuse when income is controlled for. This illustrates the problem of omitting key variables in your study, which can distort the findings.

A confounding variable is a type of control variable that obscures the effect of another variable. In such cases, the effects of a variable's impact

cannot be determined because of other influences confounding the relationship. For example, a relationship between traditional male gender role expectations and wife abuse may be confounded by a history of paternal abuse. That is, the power of traditional gender role expectations may be stronger for people who have been abused by their fathers. Several studies that look at the effect of a person's race on his or her comfort with homosexuality find that African Americans are more homophobic than whites. However, this association is misleading. When the studies add access to higher education as a variable, the impact of race disappears. That is, African Americans have attitudes about homosexuality that are similar to those of whites when they have similar educational levels.

As a result of the effect confounding variables can have on a study, researchers need to include many types of variables in their studies. A broad list of variables should include independent variables and control variables that are relevant to the study. To include all possibilities, the researcher consults the literature to identify what has already been studied. Let us examine the variables used in two separate studies. These studies used substantiated child maltreatment case files to look at the differences between families with twin children and families without twin children (DV). The variables used in the first study were:

- type of maltreatment
- number of children in the home
- annual income
- use of fertility drugs to conceive
- whether or not the children were born prematurely
- whether or not there were birth defects when the children were born
- number of adults in the home
- age of the children in the home
- square footage of the residence
- parental psychological problems
- social supports available

The variables used in the second study were:

- type of maltreatment
- number of children in the home
- annual income

Which study would you use? What makes your choice the better option?

## What Is a Measure?

A **measure** is a tool or instrument that is used to gather data. For instance, a measurement tool could be a survey (e.g., population survey), a test (e.g., IQ test), a scale that has several questions (e.g., depression scale), or a poll (e.g., opinion poll). A measure has two parts—the item (stimulus) and the response. The item is generally a statement, question, or observation that requires some type of measurable response. A measurement tool is sometimes trying to measure a concept such as alcoholism, depression, self-esteem, and marital satisfaction. Because a concept can have multiple definitions, it must be operationalized for the measure.

## Defining and Operationalizing Measures

Now let us begin looking at how we develop the items of the measure. For this, it is important to discuss how concepts are defined and operationalized. Remember, the term *operationalize* refers to how we define a concept so that it can be measured. Let us look at some examples. Alcoholism is defined as excessive use or compulsive use of alcohol in many standard dictionaries. How do we know excessive use or compulsive use when we observe drinking behavior? Whose standard do we use? If you did research on the prevalence of alcoholism at your university or college, you might find multiple conflicting statistics, depending on how you operationalized this concept. For example, in one study you may operationalize alcoholism using criteria established in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*. In another study you may simply ask whether individuals believe they are alcoholics, and for still another study, you might investigate how many individuals have been diagnosed with alcoholism by a mental health professional. As you can guess, each of these studies would lead to different findings on the prevalence of alcoholism at your university or college.

An example of an instrument for measuring alcoholism that is commonly used is the Michigan Alcoholism Screening Test (MAST). The MAST was developed by Melvin L. Seltzer (1971) to detect alcoholism. This particular scale has twenty-four questions (concepts defining alcoholism) with scores assigned for each response option (how concepts are operationalized). The total of all twenty-four scores can indicate non-alcoholism (score of 3 or less), suggestion of alcoholism (score of 4), or alcoholism (score of 5 or more). Question 8 of the instrument asks: "Have you ever attended a meeting of Alcoholics Anonymous?" If the response is yes, it is assigned

five points and indicates alcoholism according to this instrument's definition. Therefore, individuals who attended an AA meeting for reasons other than their own alcohol use, such as to support a friend, would meet the criteria for alcoholism based on how this measure was operationalized.

With this in mind, you can easily see how research can be corrupted by how a concept is operationalized. We can manipulate how we operationalize concepts such as abuse, poverty, crime, and homosexuality to fit our desired outcome or mislead readers. For example, city law enforcement might operationalize incidents of crime as all arrests or every conviction, including simple misdemeanors such as shoplifting. The findings might be used to support the need for more resources or to justify the current resources. In another study, the travel and tourism office of the same city might operationalize incidence of crime as only felony crimes against individuals. This would be used to establish a lower rate of crime for the city in an effort to attract more tourists.

Poverty provides a good example of the difficulty that can arise when we try to operationalize a concept. The battle over how to define poverty has raged for years. The issue comes down to concerns over "relative" versus "absolute" poverty. Measures of relative poverty focus on deprivation from a subjective and comparative point of view—"How poor do I feel?" or "How poor am I compared to others around me?"—whereas measures of absolute poverty focus on the amount of income an individual or family has to purchase the goods that sustain healthy life. While this definition might sound straightforward, listing all the material things people need can get very complicated. Therefore, different definitions of poverty can produce vast differences in how we operationalize poverty. The bottom line is that how we define this concept has a bearing on who will receive services.

### Levels of Measure

Four levels are used to measure variables: nominal, ordinal, interval, and ratio. The first two levels of variables are called *discrete variables*, or *categorical variables*. Both nominal-level and ordinal-level variables are discrete because they are made up of distinct separate units or categories. The last two levels of a measure are called *continuous variables* because they are made up of a large (sometimes infinite) number of units. Both interval-level and ratio-level variables fall into this category. With each subsequent level of measure, we gain the ability to more precisely measure what we

are studying. Keep in mind that we want to be as precise as possible in our measurement.

### Discrete Levels of Measure

The first level of measurement we will examine is the nominal-level variable. **Nominal-level variables** are mutually exclusive (that is, responses fit into one category and cannot be in another) and are exhaustive (no other options are available). One type of nominal-level variable is a **dichotomous variable**, for which there are only two responses to choose from (e.g., yes or no, treatment group or nontreatment group). We can measure sex as a nominal-level variable (a person is either male, female, transsexual, or hermaphrodite). We can measure attitudes at a nominal level ("Do you like research? Yes or no?"). We can measure religious affiliation as Catholic, Protestant, Jewish, Muslim, or other. One way to determine if the measure is nominal is to decide whether you can add or subtract from the measure. For instance, you cannot subtract a Catholic from a Muslim and get another religion. This highlights a problem with nominal-level variables—they are limiting. Let us say, for example, that eight people state on a survey that they are Catholic. This gives us no information about the extent of their involvement in their faith, such as church attendance, tithing, and praying. So we move up a level to ordinal-level data to gain greater detail.

With **ordinal-level variables**, in addition to being mutually exclusive and exhaustive, responses are rank ordered. For example, for the question "How much do you like research?" ordinal-level responses could range from "not at all" to "somewhat" to "very much." Notice how this provides more information than purely nominal-level responses (yes or no) for "Do you like research?" Nominal-level variables can tell us whether someone feels positively or negatively about something, but not to what degree. By using rank ordering, we are beginning to establish different levels or degrees of responses. The limitation of ordinal data, however, is that it is not precise. For instance, the difference between "somewhat" liking research and liking research "very much" can be big or small. The next level of measurement begins to achieve greater precision.

### Continuous Levels of Measure

The next level of measurement is the interval-level variable. With **interval-level variables**, items are rank ordered and each step is mutually exclusive and exhaustive, and there are equal gradations between each step. This means that the difference between the responses can be determined through addition or subtraction. An example is IQ. One person's IQ



can be 110, and another person's 140, an increase in IQ of exactly thirty points. Another example is the range between -10 and -20 degrees Fahrenheit, which is the same as that between 50 and 60 degrees, or exactly ten degrees.

The fourth level of measurement is the ratio-level variable; this is the most precise level of measurement. Ratio-level variables have all the attributes of the other three levels (items are mutually exclusive and exhaustive, they are rank ordered, and there are equal gradations between steps). The main difference between interval and ratio measurement is that interval-level variables have no absolute or fixed zero point, whereas ratio-level variables have an absolute zero point. Items measured at this level might include such things as income (you could conceivably have no income) or number of children. Notice that both income and number of children would meet all the criteria: they are mutually exclusive categories that are rank ordered and have equal gradations and an absolute zero.

Keep in mind that variables are fluid because they can be viewed and used in various ways. In table 4.1, the researcher has taken what is a ratio-level variable ("How many days in the last week have you experienced episodes of crying?") because it has an absolute zero (0-7 days) and is treating it like an ordinal variable by dividing it into categories. When this variable is treated as interval level, the results can provide more information, such as how many or what percentage reported one or two days per

**TABLE 4.1: EXAMPLE FOR EACH LEVEL OF MEASUREMENT**

ITEM (STIMULUS)	RESPONSE OPTIONS	LEVEL OF MEASUREMENT
Have you ever been treated for depression?	Yes No	Nominal
In the past month, I have thought about ending my life.	Not at all Sometimes Frequently	Ordinal
How many days in the last week have you experienced episodes of crying?	None 1-2 days 3-4 days 5-6 days Daily	Interval
How many times have you attempted suicide?	(Respondents enter a number.)	Ratio

week as opposed to three or four per week, five or six days per week, daily, or none.

In the following examples, indicate what level of measurement is being used:

- A student is polling other students on campus about their position on abortion. Students have the option of defining their position on the issue as "support abortion" or "do not support abortion."
- A substance abuse counselor wants to know how satisfied consumers are with the on-site Alcoholics Anonymous meetings. She administers a survey that asks respondents to indicate if they are "very satisfied," "somewhat satisfied," or "not satisfied."
- The supervisor of an after-school program needs to know in what grade each student is currently enrolled.

### Standardized Measures

There are multiple standardized measurement instruments in circulation today. A standardized measure is one that has been given to enough people that we can compare one person's scores to those of other test takers. For example, you have probably taken the ACT or a similar exam. This test has been normed or standardized so that your scores could be compared with those of other test takers. A similar example would be tests of intelligence quotient (IQ tests). Most people who take this test score in the range of average intelligence (around 100). Some compilations of measures are described at the end of this chapter.

Standardized measures have limitations. To illustrate some important limitations, let us return to the example of the Michigan Alcoholism Screening Test. The MAST was normed on multiple groups: 116 hospitalized alcoholics, 99 people arrested for drunk driving, 98 people under review for revocation of their driver's licenses because of excess accidents and moving violations, and 103 controls (individuals who functioned as part of a control group). The groups were largely made up of white males between the ages of twenty-five and forty-four years. What are potential problems that might arise if the MAST is administered in the following situations?

- You are working with members of a gang who are in high school. Drinking alcohol appears to be an important part of their culture.

- You are working with parents referred by Child Protective Services for alcohol abuse. The majority of these parents are women.
- You are working with Hispanic immigrants referred through the court system for charges of driving while intoxicated.

Standardized measures such as the MAST are used universally, regardless of the population for which they have been normed. Researchers should be informed of how standardized measures are normed so that they can avoid placing individuals into stereotypical categories that do not take into consideration issues of diversity such as culture, sex, age, and nationality.

### Reliability and Validity in Measurement

Two of the factors that contribute to the credibility of a research study are the measurement's reliability and validity.

**Reliability** is the word used to describe the stability and consistency of a measurement. For instance, a tape measure is a highly reliable measuring instrument because it does not change over time (i.e., it is stable), and it measures everything according to the same standards (i.e., it is consistent). There are four major categories of reliability for most instruments: test-retest, equivalent form, internal consistency, and interobserver reliability.

**Test-retest reliability** has to do with the consistency of your measure from one time to the next. When a researcher administers the same measurement tool multiple times to the same group following the same research procedures, does he or she obtain consistent results (assuming that there has been no change in what is being measured)? If so, the measure has test-retest reliability.

**Equivalent form reliability** is concerned with consistency between two versions of a measure. If a researcher wanted to develop a new measurement tool for anxiety, he or she could administer the new measurement tool alongside a more traditional measurement tool to the same group. If consistent results were found, it could be argued that the new measurement tool had reliability for measuring the concept of anxiety. Another option is to create two new instruments that measure the same concept. This requires a long list of variables so that the concept can be divided between the two instruments. Again, both instruments would be administered at the same time to the same group, and then the researcher would compare the scores from both instruments to determine if they were equivalent.

Internal consistency is the consistency among the responses to the items in a measure. It is the extent to which responses to items measuring the same concept are associated with each other. This form of reliability is examined when a single measurement instrument is administered to a group of people on one occasion. Tests of internal consistency estimate reliability by grouping items in an instrument that measure the same concept. For example, you could write two sets of three questions each that measure the same concept (e.g., social isolation) and if the responses to those two groups of three questions were consistent, you would know that your instrument is reliably measuring that concept. For instance, if we were measuring the concept of social isolation, we might ask respondents if the statements "I feel lonely at times" and "I never feel lonely" accurately describe their experiences, and we would expect these items to have opposite answers.

Another practice used to determine internal consistency is called split-half reliability, which involves randomly dividing the items related to a concept into two groups. Then the scores of the two groups are compared to determine if they are measuring the same concept. In this way, you are not developing two surveys, as required by the equivalent form method. The more items that are included and the stronger the consistency between the two groups of scores, the greater the reliability is.

Interobserver reliability must be measured when more than one observer uses the same instrument to rate the same person, place, or event. If different observers or interviewers use the same instrument to score the same thing, their scores should match. The more similar the ratings, the more reliable the findings are. For instance, four observers are using an observational assessment tool to measure the quality of interactions between a mother and her child on a playground. The rating scale of the tool ranges from 0 to 10, and two observers rate the mother as a 4, while another observer rates the mother as a 2, and the final observer rates the mother as a 6. These scores are fairly similar to each other. This measure would be less reliable than if the ratings were more widely spread, for instance, scores of 2, 3, 6, and 10.

An alternative version of this type of reliability is called intraobserver reliability. This means that there is one observer rating a person, place, or event two or more times. As with interobserver reliability, the findings are compared to ensure that the measurement is consistently getting similar scores each time.

Validity is a term that is used to describe how much a measurement tool (such as a scale, survey, poll, or test) measures what it is meant to measure. It is the match between how a concept is conceptualized (defined)

and how it is operationalized (measured). For example, if we are using a scale that measures depression, we would expect the scale to ask questions about changes in the person's eating and sleeping habits (either eating or sleeping more or less), number of suicidal ideations, and the like. If these were the questions that were asked, we might say the scale had validity. Conversely, the measurement would not focus on other psychological traits such as anxiety, poor body image, or attention deficits. A few years ago, a national fast-food chain bragged that it had the best french fries in the world. This assertion was based on their claim to have sold the most french fries globally. How is the validity of this claim flawed? In this example, the focus was on the number of fries as opposed to the quality of taste. Think for a moment about the issue of availability. This fast-food chain has a global presence; the more stores there are, the more fries they sell. But this does not mean that the claim that the chain's french fries are better is valid.

The fact that a measurement instrument is reliable (it is consistently measuring the same thing over and over) does not mean that it is valid. For example, let us say that you want to measure your participants' level of depression. You select a measure that has been normed and is standardized. However, the instrument you select is a scale that measures how a person feels about him- or herself (a measure of self-esteem). Even though you are consistently measuring your participants' self-esteem, you may not be measuring their level of depression. This is an issue of validity.

Let us return again to the MAST instrument. The validity of the measure was established in the following ways:

- The MAST was able to classify respondents as alcoholic or non-alcoholic.
- Only 15 out of 526 people originally classified as non-alcoholic by the MAST were found to be alcoholic.
- The MAST correctly identified 92 percent of the 99 respondents hospitalized for severe alcohol problems.

Consider the following questions:

1. The MAST classified respondents as either non-alcoholic or alcoholic. How do we know these classifications are correct? (Think about what you have learned about operationalizing and defining measures.)
2. Should the study report how many people who were classified as alcoholic were actually non-alcoholic?

3. Does administering a survey to hospitalized alcoholics influence the findings?

So how do we establish validity? There are several types of validity: face validity, content validity, criterion-related validity, concurrent validity, and construct validity.

### Face Validity

Face validity refers to whether a measure seems to make sense (be valid) at a glance. When a student asks another student to look over a paper to see if his or her answers appear to be correct, the student is requesting a review of the face value of the paper. However, when the paper is evaluated by the standards of the instructor and the overall performance of the other students on their papers (the norm), the comments the student receives from the instructor may differ from those he received from the student who looked only at the face value of the paper. Therefore, caution must be taken when one is determining the validity of research based on the face value of what is presented. In fact, many researchers do not consider face validity a useful form of validity, and face validity should never be the only form of validity used to validate a measure.

The following are some issues that you can look for when you are examining the face value of a measurement tool.

1. Forced generalities. These are questions that force a respondent to make generalizations. An example would be "Do you trust people in your community?" A yes answer implies blanket trust in everyone in the community.
2. Inapplicable items. Some items do not apply to all people or their situations. For instance, the question "Are you close to your father?" implies that everyone has a father.
3. Double-barreled questions. A variable that asks two questions but only allows for one reply is a double-barreled question—for example, "On a scale from 1 to 5, how would you rate your satisfaction with your supervisor and colleagues?" In this situation, respondents may be happy with their supervisor, but not with their colleagues.
4. Unclear items. An unclear item may lead to responses that are unintended because the question was misunderstood. An example would be asking: "Do you visit your advisor regularly?" The interpretation of "regularly" is different for each respondent.

5. Leading questions. A leading question is one that tries to lead a respondent to a particular response, for instance, "With the rising popularity of violent video games, do you think there should be mandatory ratings for children?"
6. Overdemanding recall questions. "Have you been sent to the principal's office in the last week?" is a question that demands less recall than "How many times have you been sent to the principal's office in your lifetime?"

### Content Validity

Content validity refers to how well the items in a measurement represent the concept that is being measured. For example, going back to the issue of conceptualizing child maltreatment, by reviewing the child maltreatment literature, a researcher developing a tool to measure child maltreatment would find that the measure would need to include the commonly identified types of both abuse and neglect.

As with face validity, there can be difficulties in relying solely on content validity. Experts may disagree concerning the range of content provided in a measure. For example, a scale measuring obsessive compulsive disorder (OCD) that asked, "Do you have great difficulty discarding things even when they have no practical value?" would pose problems if experts do not agree that this item is representative of the concept of OCD. In this instance, while hoarding behavior *can* be related to OCD, many people without OCD hang on to items that they never use.

### Criterion-Related Validity

Criterion-related validity refers to a measure's ability to make accurate predictions and is also referred to as predictive validity. The name *predictive* comes from the fact that this form of validity is derived from how well the measure predicts an outside criterion. For instance, how well SAT scores predict college grades is an indication of the SAT's criterion validity. As another example, a scale measuring students' satisfaction with their educational experience at a university could be compared to how many incoming students stay at that university (graduation rate). In this case, student satisfaction predicts the criterion (graduation). This type of validity is difficult to establish because researchers may not be able to gain access to the criterion, such as the specific course grades or graduation rates of the sample.

### Concurrent Validity

Concurrent validity refers to how well a measure correlates with some other measure of the same variable that is believed to be valid. For instance, if a researcher designed a survey to measure depression, he or she could compare it to another measure of depression to see if they predict the same outcome. For example, if an instrument measuring the concept of job satisfaction gives results that are similar to those given by a job satisfaction instrument that has already been validated, the new measurement has concurrent validity. Therefore, when concurrent validity is being measured, the two measures are taken at the same time. This is different from measuring predictive validity, where one measure (e.g., a measure of job satisfaction) is meant to predict responses to a measure that is administered at a later time (e.g., a measure of job retention).

### Construct Validity

Construct validity refers to the extent to which the items of an instrument accurately sample a construct. A construct is the concept or the characteristic that an instrument is designed to measure. Whereas measures of content validity ask whether an instrument measures the full range of possibilities within a concept (e.g., alcoholism), construct validity is the degree to which an instrument actually reflects the construct being measured. In other words, with content validity we are determining whether the instrument includes the items (content) that can accurately operationalize a concept, and with construct validity we are determining whether the overall instrument was constructed to measure a single concept such as alcoholism, and not OCD or social isolation. This type of validation is commonly used in social research when there is no existing criterion for validation purposes. There are two types of construct validity: convergent validity and discriminate validity.

Convergent validity refers to how well the measures of a construct (e.g., depression or alcoholism) that you expect to be related to each other are, indeed, found to correspond to each other (measure the same construct). Conversely, discriminate validity refers to the degree to which the measures of a construct that you would not expect to be related are indeed measuring different constructs.

### Applied Learning Activities

#### Activity #1

Identify the independent variable and the dependent variable:



1. Amount of time studying and scores on a final exam
2. The number of divorces a mother has had and her children's fear of intimacy in adult relationships
3. A person's sex and fear of intimacy in adult relationships
4. Number of hours a child spends playing violent video games and the child's aggression scores on a child behavior scale
5. Scores on SATs and grade point average in the freshman year of college
6. Use of Ritalin or Cylert and amphetamine usage during adolescence

#### Activity #2

Sandy is a social worker at a nursing home. Sandy notices that residents of the nursing home who are more physically active seem to be less depressed, have more energy, and generally seem to be healthier than the more sedentary residents. Sandy reviews the literature and finds a relationship between physical activity and depression in the elderly. Based on her literature review, Sandy develops the following research hypothesis: Residents of the nursing home who exercise a minimum of twenty minutes a day, three times a week, will have lower levels of depression than those who don't. Sandy develops the following research design to test her hypothesis:

- She recruits volunteers from the residents to participate in an exercise class. The class meets three times a week for thirty minutes. A total of twenty residents volunteer to participate in the exercise class.
  - She asks residents who don't wish to participate in the exercise class to act as a comparison group by taking a pre- and posttest. She is able to obtain a comparison group of nineteen residents who will take the pre- and posttests (but are not willing to participate in the exercise classes).
  - She gives members of both groups a standardized depression inventory at the beginning of the study (pretest).
  - After four weeks, a total of fifteen people complete the exercise classes and take the posttest. Sixteen of the original nineteen members of the comparison group complete the posttest. Sandy compares the scores of the two groups.
1. What are the strengths of Sandy's design?
  2. What are the weaknesses of this study?

3. How can you determine measurement validity?
4. How can we determine if the measure is reliable?

### Key Points

- Variables can be categorized into three groups: independent variables, dependent variables, and control variables.
- The dependent variable is predicted by another variable, or is said to depend on the independent variable.
- The independent variable is often thought of as a variable that is manipulated by the researcher.
- Conceptualizing a variable refers to how we translate an idea or abstract theory into variables that can be used to test hypotheses or make sense of observations.
- A control variable is any variable that the researcher wants to hold constant (control for).
- A confounding variable obscures the effect of another variable.
- A measure is a tool or instrument that is used to gather data.
- There are four levels of variables: nominal-level variables, ordinal-level variables, interval-level variables, and ratio-level variables.
- The term *operationalize* refers to how we define a concept so that it can be measured.
- Reliability refers to the ability of a measure to remain stable and consistent over time.
- Validity is a term that describes how much the instrument measures what it is meant to measure.
- There are several types of validity: face validity, content validity, criterion-related validity, concurrent validity, and construct validity.

### Additional Resources

Fischer, J., & Cochran, K. (2007). *Measures for clinical practice: A sourcebook* (4th ed.). New York: Oxford University Press.

This is a two-volume set of books that compiles multiple measures devoted to couples, families, and children. A description of the purpose, norms, scoring, reliability, validity, primary reference, and availability is provided for each measure.

#### Mental Measurements Yearbooks

The University of Nebraska Press and Rutgers University Press both produce a yearbook of mental measurements.

Robinson, J., Shaver, P., & Wrightsman, L. (Eds.). (1999). *Measures of political attitudes*. San Diego, CA: Academic Press.

This book offers a compilation of measurement scales on issues such as racial attitudes, trust in the government, and political participation.

