Chapter 4: Defining and Measuring Variables

A. LEARNING OUTCOMES. *After studying this chapter students should be able to:*

- Distinguish between qualitative and quantitative, discrete and continuous, and independent and dependent variables.
- Describe what constructs, mediator variables, and moderator variables are.
- Explain why it is important for scientists to be explicit about how they conceptually define their variables.
- Discuss what it means to operationally define a variable, and describe some examples of how a construct can be operationally defined.
- Describe the basic characteristics of nominal, ordinal, interval, and ratio scales.
- Provide examples of measures that reflect these four measurement scales.
- Describe what is meant by the accuracy and reliability of a measure.
- Explain how test–retest reliability and internal-consistency reliability can be determined.
- Explain how validity differs from reliability, and describe the various types of validity.

B. KEYWORDS

Accuracy
Construct
Construct validity
Content validity
Continuous variable
Criterion validity
Dependent variable
Discrete variable
Face validity
Independent variable
Internal-consistency reliability
Interval scale
Measurement
Mediator variable
Moderator variable
Nominal scale
Operational definition
Ordinal scale
Qualitative variable
Quantitative variable
Random measurement error
Ratio scale
Reliability
Scales of measurement
Situational variable
Subject variable
Systematic error (bias)
Test–retest reliability
Validity
Variable

C. BRIEF CHAPTER OUTLINE

I. Types of Variables

A. Qualitative and Quantitative variables
B. Discrete and Continuous Variables
C. Independent and Dependent Variables
D. Constructs
E. Mediator and Moderator Variables

II. Defining Variables

A. Conceptual Definitions
B. Operational Definitions
C. Operational Definitions in Everyday Life

III. Scales of Measurement

A. Nominal
B. Ordinal
C. Interval
D. Ratio

IV. Measurement Accuracy, Reliability, and Validity

A. Accuracy of Measurement
B. Reliability of Measurement
C. Validity of Measurement
D. Some Final Thoughts

D. EXTENDED CHAPTER OUTLINE

**Much of this summary is taken verbatim from the text.*

Introduction

This chapter discusses what variables are, how they are defined and measured, and how their accuracy can be assessed in research.

Part I: Types of variables

A *variable* is any "thing" that can take on two or more values. In contrast, a *constant* is any "thing" that cannot change (i.e., the mathematical value *pi*). Variables can be:

A. **Qualitative or quantitative.** Whereas qualitative variables differ in type, quantitative variables differ in amount. In other words, qualitative = what kind, and quantitative = how much.
B. **Discrete or continuous.** Discrete variables are those in which their associated measures do not have an infinite number of values between them. The number of children a person has is a discrete variable; a person can have one child or two children, but not 1.5 children. In contrast, continuous variables do take on an infinite number of values.

C. **Independent or dependent.** Independent variables are the presumed causal factor in experiments, whereas dependent variables are the presumed effect. Independent variables can be further classified into *situational variables* or *subject variables.* Situational variables are those that are under the direct control and manipulation of the experimenter (drug, placebo). Subject variables are those that naturally exist in participants (age, gender).
D. **Constructs** are underlying, hypothetical characteristics or processes that are not directly observed but instead are inferred from measureable behavior or outcomes (i.e., depression, self-esteem, intelligence). One cannot see constructs but can infer them based on behavior.
E. **Mediator variables** are those variables that provide a causal link in the sequence between the independent and dependent variables and help explain "why" a behavior occurs. **Moderator variables** are those that alter the strength or direction of the relation between an independent and dependent variable and help explain "when" and "for whom" a behavior occurs.

Part II: Defining Variables

Our general understanding of what something means is a **conceptual definition.** For example, we all understand what stress is but when we are asked to define it precisely our answers will differ to some degree. In research, scientists minimize the ambiguity of what something means by using **operational definitions.** Operational definitions define a variable in terms of the procedures used to measure or manipulate it, which enables one to precisely communicate with others about what something is. An operational definition of stress might be one's score on a Stress Inventory in which the range of scores is 5 (low stress) to 25 (high stress). The term *operational definition* may be new term to some, but people use operational definitions every day. Job descriptions and GPA are two examples of how we operationalize our specific duties at work and our performance in school.

Part III: Scales of Measurement

Measurement is the process of systematically assigning values to represent attributes of organisms, objects, or events. There are four **scales of measurement**, each of which defines rules for assigning scale values to measurements.

A. **Nominal scales** of measure represent only qualitative differences of the attribute of interest. That is, they categorize data based on type (i.e., political party). Nominal scales create a set of labels for categories (Democrat, Republican, Independent) that are mutually exclusive and to which participants are assigned. Nominal scales are the weakest form of measurement because they assume equality among members of a particular category and differences between categories. In addition, the numerical value assigned to each category is arbitrary.
B. The values in **Ordinal Scales** represent relative differences in the amount of some attribute such that the value assigned to someone or something reflects its rank among the rest of the data. An example is *U.S. News & Report*'s "America's Top College." Unlike nominal scales, ordinal scales indicate how much scores differ from one another.
C. **Interval scales** also reflect quantitative differences among data, but unlike ordinal scales they assume equal differences in the amount of the attribute being measured. Temperature measured on the Fahrenheit scale is an interval measure because each incremental change in temperature (70, 71, 72, etc.) is the same.

D. **Ratio scales** are the most sophisticated form of measurement. There are equal distances between values on the scale and the scale has a true zero (0) point. That said, a score of 0 indicates a true absence of the attribute in question. This enables one to create meaningful ratio, such as "10 is twice as much as 5." Interval scales cannot make these statements because their 0 point is arbitrary.

Part IV: Measurement Accuracy, Reliability, and Validity

A. The **accuracy of a measure** represents the degree to which the measure yields results that agree with a known standard. **Systematic error (bias)** is a constant amount of error between the true value and the measurement of it. Scientists calibrate instruments to known standards in order to reduce bias in measurement. Because some things in psychology cannot be calibrated (e.g., personality, I.Q.) psychologists rely heavily on reliability and validity of measurement.

B. **Reliability** refers to the consistency in measurement. If you were to weigh yourself five times over a period of five minutes, your bathroom scale would yield a relatively consistent (within +/– 0.1 lb) measure. The random variation that may occur across the five measures reflects **random measurement error:** fluctuations in the measuring situation that cause the obtained score to deviate from its true score. Not that reliability does *not* mean a measure is accurate. You could weigh in at 150 while your true weight is 155.
   a. When a measure is administered to the same participant on two or more occasions under equivalent conditions, reliability of that measure can be determined through **test–retest reliability.** Another way to assess reliability is through **internal-consistency reliability.** This is the consistency of a measure within itself and it is typically determined by splitting an instrument into two halves and examining how each half covaries with the other.

C. **Validity** in a measure refers to whether a measure assesses what it claims to assess. For example, the Beck Depression Inventory is a valid assessment of depression if it truly measures depression. There are several ways to examine validity.
   a. **Face validity** is the degree to which the items on a measure appear to reasonably reflect what it is that the assessment should measure. In other words, "It looks good."
   b. **Content validity** is the degree to which the items on a measure adequately represent the entire range or set of items that could have been appropriately included.
   c. **Criterion validity** is the ability of a measure to predict an outcome. For example, performance on the SAT should predict college GPA.
   d. **Construct validity** is when a measure truly assesses the construct it claims to measure. Construct validity can be established through:
      i. **convergent validity —**scores on a measure should correlate with scores on a similar measure, or
      ii. **divergent validity —**scores on a measure should not correlate with scores on measures of different constructs.

E. LECTURE AND CLASSROOM ENHANCEMENTS

PART I: Types of Variables

A. Lecture/Discussion Topics

- **Independent variables (IV), dependent variables (DV)…what's the big difference?** Students often have a conceptual understanding of the difference between independent and dependent variables but they interchange the terms (i.e., they describe an IV but call it a DV, and vice versa). The strategy I use to help students correctly name each variable is as follows: "Independent variables are those things a researcher manipulates. I am a researcher. In my experiments **i**ndependent **v**ariables are things that ***I v***ary. By making the connection between the phrase "I vary" and the abbreviation "IV," for independent variable, I've found that students have been able to use the terminology correctly. Likewise, I tell my students that in psychology we measure behavior and that the **d**ependent **v**ariable is the thing that is measured and that the behavior a person produces is *dependent* on which treatment condition he or she was exposed to.
- **Variables and constants.** A variable, by definition, is anything that can take on two or more values. In contrast, a constant is something that only has one value at any given time. Why do psychologists measure variables and not constants? Because if a constant exists in one case it exists in all cases. In other words, if a particular trait or behavior is a constant there would be no need to examine it after it's been identified because it would exist in the same way in every single human being. So what things are constant? Very few true constants actually exist. Time is both a variable *and* a constant. For example, although the time of day changes second to second, the number of seconds in a minute does not change. So what things are true constants? There are several in mathematics. *Pi* (3.14), for example, is a constant that when multiplied by the diameter of any size circle will equal the circle's circumference.
- **How to spot a qualitative variable.** All behaviors, characteristics, and traits that are measured at a nominal scale are qualitative variables, or are at least defined qualitatively for that particular purpose. So when you are asked if a dependent variable is qualitative or quantitative, you should first determine what scale that something is being measured at. If it's nominal, then it's absolutely qualitative.
- **An experiment's purpose determines a variable's classification.** Variables are usually described as mutually exclusive. While this is true for the *type* of variable (IV, DV, moderator, extraneous), it is not necessarily true for the variable itself. For example, anxiety can be an independent variable manipulated to create two treatment conditions in one experiment, and can be what is measured (the DV) in a different experiment. In yet another study, anxiety could be a confounding variable that prevents the researcher from making cause and effect statements regarding learning style and test performance. The point is that many things can be either an IV or DV, depending on the purpose of the study. (See the exercise below to further illustrate this point.)

B. Classroom Exercise

- **An experiment's purpose determines a variable's classification.** To understand how any "thing" can be an independent variable or dependent variable depending on the purpose of the study, provide students with the following pairs of variables. Have them create two testable hypotheses for each pair—one using variable A as the IV and B as the DV, and the other with variable B as the IV and A as the DV.

Variable A	Variable B
Sleep	Anxiety
Workload	Stress
Music Genre	Arousal
Risk taking	Alcohol

C. Web Resources

- **One of the worst driving distractions on the road: Your kids.** http://abcnews.go.com/US/worst-driving-distractions-road-kids/story?id=18684711#.UY8DU8q3Nec
- **Variables explained**. This webpage summarizes the difference between variables and constants, as well as quantitative and qualitative variables, the levels of quantitative variables that exist, and independent, dependent, and moderator variables. http://www.gifted.uconn.edu/siegle/research/variables/variablenotes.htm

D. Additional References

- **Studies that measure both qualitative and quantitative variables.**

De Witte, H., De Cuyper, N., Handaja, Y., Sverke, M., Näswall, K., & Hellgren, J. (2010). Associations between quantitative and qualitative job insecurity and well-being. *International Studies of Management and Organization,* 40(1), 40–56.

Johnston, O., Reilly, J., & Kremer, J. (2011). Excessive exercise: From quantitative categorization to a qualitative continuum approach. *European Eating Disorders Review*, 19(3), 237–248.

PART II: Defining Variables

A. Lecture/Discussion Topics

- **Illustrating the difference between conceptual and operational definitions.** An increasing body of literature indicates that physical activity is very beneficial to one's physical and mental health. But what, exactly, constitutes physical activity? Is running several miles at least three times per week necessary to reap the benefits of exercise or is simply taking the stairs rather than the elevator sufficient? Although running and taking the stairs are both forms of exercise, they are qualitatively different and may have different effects on one's health. To demonstrate the importance of defining variables, have the class help you create a simple experiment to determine whether exercise improves cognitive function. The simple answer is to create two treatment conditions: one with those who exercise and one with those who do not, and then measure cognitive function. However, what type of exercise and how much of it must one perform in order for an effect on cognition to be observed? Walk across the classroom, asking students: "is this sufficient exercise?" It *is* exercise, right? Although it's unlikely that a psychologist would have his "walkers" only travel the distance equivalent to the classroom's width, if one doesn't specify exactly what constitutes walking, there is way too much room for interpretation. Research studies should be designed such that all variables (those manipulated and those measured) are clearly and precisely defined.

B. Classroom Exercise

- **The value of operationism in research.**
  - To illustrate the importance of operational definitions and why they are superior to conceptual definitions, ask the class to keep count of the aggressive acts exhibited in this *Tom & Jerry* video. After the video have students report the number of aggressive acts in it. The number should vary significantly. Some students may have only counted Tom's aggressive acts, others only Jerry's aggressive acts. Some may count an aggressive act only if it led to harm, while others count all acts whether they ended up inflicting harm or not. The point is that while we all know what aggression is, we need a clear set of operations to specify exactly what is meant by aggression. http://www.youtube.com/watch?v=RzEZ0-c_JZM
  - Have students develop operational definitions for the following terms: memory, intelligence, exercise, and sleep deprivation. Then ask them to present their operational definitions to the class. Write the operational definitions on the board to illustrate how one person's definition of something is very different from another's. When students see that multiple definitions were generated for the same term, they will appreciate the need for operationism in science.

C. Web Resource

- **Operational Definitions.**
  - An Xtranormal clip on the operational definition of "satisfaction": http://www.youtube.com/watch?v=37dLMgWPAtM
  - Operational definitions clarified: http://www.ablongman.com/graziano6e/text_site/MATERIAL/opdefsex.htm

D. Additional References

Ennis, R. H. (1964). Operational definitions. *American Education Research Journal, 1,* 183–201.

Nordgarrd, J. & Parnas, J. (2013). A haunting that never stops: Psychiatry's problem of description. *Acta Psychiatrica Scandinavica, 127,* 434–435. *(An interesting paper on the need to operationalize the criteria for psychiatric disorders.)*

PART III: Scales of Measurement

A. Lecture/Discussion Topics

- **There are many ways to measure a single variable.** Every attribute we seek to examine in research can be quantified at least nominally, and many attributes could be quantified at the ordinal, interval, and ratio measure as well. Time, for example, is often provided as the classic example of a ratio measure. We could measure reaction time by recording the number of seconds it takes someone to respond to a visual stimulus. However, instead of measuring time in seconds, we could measure time by rank, ordering one person's response time so that there is a first fastest, second fastest, third fastest, and so on. We could also measure reaction time nominally by grouping people into the categories of "reacted" and "did not react." Because many attributes can be measured in a variety of ways, researchers must carefully select the scale of measurement that will provide the best, richest information about that attribute. Ratio scales do just that. However, one should not select an attribute based on its potential to be measured on a ratio scale. Some attributes are inherently nominal (political party) and, as such, a nominal scale is the only—and best way—to measure that attribute.
- **Is gender really a nominal variable?** Most research requires participants to provide a variety of demographic information including their gender. In the United States gender is typically considered in terms of being male or female. Is male-female a clear dichotomy? In Japan gender is seen as… What about transgender and transsexual individuals? While the number of individuals who

B. Classroom Exercises

- **Quantifying the same behavior with different levels of measurement.** Provide students with several attributes (i.e., memory, intelligence, exercise, and sleep deprivation) and have them identify which scales of measurement could be used for each. Encourage students to identify a measurement for each attribute at every level (nominal, ordinal, interval, and ratio). For example, for memory a nominal measure might be grouping people based on whether they did or did not memorize a set of words; an ordinal measure might be rank-ordering people based on the number of correct items recalled; a ratio measure could be the number of words one correctly recalls.

C. Web Resources

- **Scales of measurement.** This webpage has a brief video (PPT video, about 8 minutes) about scales of measurement as well as a written summary of each scale.
  http://stattrek.com/statistics/measurement-scales.aspx

D. Additional References

- **Measurement is tied to the way in which variables are defined.** Phillips, S. P. (2005). Defining and measuring gender: A social determinant of health whose time has come. *International Journal of Equity in Health, 4,* 11.

PART IV: MEASUREMENT, ACCURACY, RELIABILITY, AND VALIDITY

A. Lecture/Discussion Topics

- **Using Pearson product-moment correlation to determine reliability.** Reliability has to do with consistency in a measure. If an instrument consistently yields the same measure, it is said to be a reliable assessment of that construct. Although we can take several measures and make a general statement of the instrument's reliability, you can perform statistical analyses to make a more definitive conclusion. To determine the reliability of a new instrument, measures can be obtained from several people at time 1 and then again at time 2. The bivariate distribution of data can then be examined via correlation. Correlational analyses provide a numerical index of relatedness between two variables, in this case measurement 1 and measurement 2, such that the closer the value is to 1.00, the more reliable the instrument is.
- **The reliability and validity of psychological surveys.** Psychologists often use surveys to measure characteristics such as personality, mood, and motivation. For any construct of interest there are a variety of measures. With regard to depression, for example, there is the Beck Depression Inventory, Hamilton Depression Scale, Center for Epidemiologic Studies Depression Scale, Multiscore Depression Inventory, and Zung Self-Rated Depression Scale. If one needs to measure depression, then, which instrument should he use? There are a number of factors that contribute to which one should be selected, but the most important factor should be whether it is a good measure. In other words, is the instrument a reliable and valid measure of the construct of interest? Every good instrument should have undergone a rigorous assessment of its reliability and validity, and the results of the assessment should be demonstrated in a peer-reviewed published summary of the instrument.

B. Classroom Exercise

- **Determining reliability and validity.** To give students practice at assessing the reliability and validity of psychological instruments, find a source article (the paper that describes the development and psychometric results) of three instruments, each of which claim to measure the same construct. Have the students work in small groups to review the source articles and decide which instrument they would use if they had to decide among the three.

C. Web Resource

- **On reliability and validity.**
  - http://psychcentral.com/blog/archives/2011/10/16/scientific-measures-reliability-and-validity/
  - http://psychology.ucdavis.edu/sommerb/sommerdemo/stantests/test_rel.htm

D. Additional References

- **On reliability and validity.**
  Sechrest, L. (1984). Reliability and validity. *Research methods in clinical psychology*, 24–54.