Chapter 5: Measurement

The heart of science is measurement.
-Erik Brynjolfsson

Researchers Tara MacDonald and Alanna Martineau were interested in the effect of female college students' moods on their intentions to have unprotected sexual intercourse (MacDonald & Martineau, 2002). [1] In a carefully designed empirical study, they found that being in a negative mood increased intentions to have unprotected sex—but only for students who were low in self-esteem.

Although there are many challenges involved in conducting a study like this, one of the primary ones is the measurement of the relevant variables. In this study, the researchers needed to know whether each of their participants had high or low self-esteem, which of course required measuring their self-esteem. They also needed to be sure that their attempt to put people into a negative mood (by having them think negative thoughts) was successful, which required measuring their moods. Finally, they needed to see whether self-esteem and mood were related to participants' intentions to have unprotected sexual intercourse, which required measuring these intentions.

Clinicians also use surveys and other measurements, There are thousands available to measure everything from anxiety to marital satisfaction. Hundreds are readily available to the Human Service worker that are in the public domain. One popular one is the Rosenberg Self-Esteem Scale.

Do You Feel You Are a Person of Worth?

The Rosenberg Self-Esteem Scale (Rosenberg, 1989) [2] is one of the most common measures of self-esteem and the one that MacDonald and Martineau used in their study.

Participants respond to each of the 10 items that follow with a rating on a 4-point Likert scale: Strongly Agree, Agree, Disagree, Strongly Disagree. Score Items 1, 2, 4, 6, and 7 by assigning 3 points for each Strongly Agree response, 2 for each Agree, 1 for each Disagree, and 0 for each Strongly Disagree. Reverse the scoring for Items 3, 5, 8, 9, and 10 by assigning 0 points for each Strongly Agree, 1 point for each Agree, and so on. The overall score is the total number of points. The actual form is included in the appendix, and the questions used for the scale are:

- I feel that I'm a person of worth, at least on an equal plane with others.
- I feel that I have a number of good qualities.
- All in all, I am inclined to feel that I am a failure.
- I am able to do things as well as most other people.
- I feel I do not have much to be proud of.
- I take a positive attitude toward myself.
- On the whole, I am satisfied with myself.
- I wish I could have more respect for myself.
- I certainly feel useless at times.
- At times, I think I am no good at all.

Likert scales are often used on surveys. Below is a five point Likert scale which allows the respondent to take a neutral stance. There are four point scales that force the respondent to agree or disagree one way or the other. The middle option is missing. Four and five point scales are the most common but occasionally we use surveys with more or less options available.

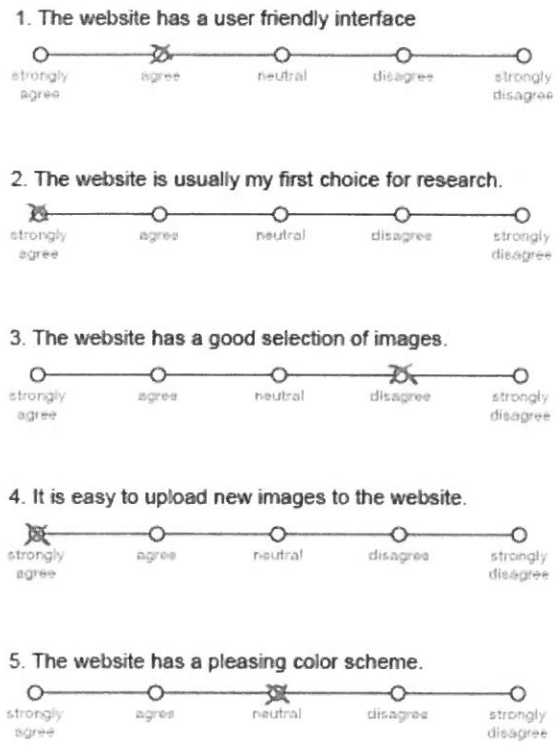

FIGURE 3: FOUR POINT LIKERT SCALE FROM:
HTTPS://UPLOAD.WIKIMEDIA.ORG/WIKIPEDIA/COMMONS/D/D7/EXAMPLE_LIKERT_SCALE.JPG

[1] MacDonald, T. K., & Martineau, A. M. (2002). Self-esteem, mood, and intentions to use condoms: When does low self-esteem lead to risky health behaviors? *Journal of Experimental Social Psychology, 38*, 299–306.

[2] Rosenberg, M. (1989). *Society and the adolescent self-image* (rev. ed.). Middletown, CT: Wesleyan University Press.

5.1 Understanding Clinical Measurement

LEARNING OBJECTIVES

- Define measurement and give several examples of measurement relevant to Human Services practice.
- Explain what a construct is and give several examples.
- Distinguish conceptual from operational definitions, give examples of each, and create simple operational definitions.
- Distinguish the four levels of measurement, give examples of each, and explain why this distinction is important.

What Is Measurement?

Measurement is applying a score to a characteristic. In psychology, social work, and generally in the social sciences, **measurement** is very often the assignment of scores to individuals so that the scores represent some characteristic of the individuals.

This very general definition is consistent with the kinds of measurement that everyone is familiar with—for example, weighing oneself by stepping onto a bathroom scale, or checking the internal temperature of a roasting turkey by inserting a meat thermometer. It is also consistent with measurement throughout the sciences. In physics, for example, one might measure the potential energy of an object in Earth's gravitational field by finding its mass and height (which of course requires measuring those variables) and then multiplying them together along with the gravitational acceleration of Earth (9.8 m/s^2). The result of this procedure is a score that represents the object's potential energy.

Of course, this general definition of measurement is consistent with measurement in the Human Services as well (in psychological measurement is often referred to as **psychometrics**). Imagine, for example, that a cognitive psychologist wants to measure a person's working memory capacity—his or her ability to hold in mind and think about several pieces of information all at the same time. To do this, she might use a backward digit span task, where she reads a list of two digits to the person and asks him or her to repeat them in reverse order. She then repeats this several times, increasing the length of the list by one digit each time, until the person makes an error. The length of the longest list for which the person responds correctly is the score and represents his or her working memory capacity. Or imagine a clinical psychologist who is interested in how depressed a person is. He administers the Beck Depression Inventory, which is a 21-item self-report questionnaire in which the person rates the extent to which he or she has felt sad, lost energy, and experienced other symptoms of depression over the past 2 weeks. The sum of these 21 ratings is the score and represents his or her current level of depression.

The important point here is that measurement does not necessarily require any particular instruments or procedures. It does not require placing individuals or objects on bathroom scales, holding rulers up to them, or inserting thermometers into them. What it does require is some systematic procedure for assigning scores to individuals or objects so that those scores represent the characteristic of interest.

Constructs

Many variables used by Human Services practitioners, and researchers, are straightforward and simple to measure. These include sex, age, height, weight, and birth order. You can almost always tell whether someone is male or female just by looking. You can ask people how old they are and be reasonably sure that they know and will tell you. Although people might not know or want to tell you how much they weigh, you can have them step onto a bathroom scale. Other variables studied by psychologists—perhaps the majority—are not so straightforward or simple to measure. We cannot accurately assess people's level of intelligence by looking at them, and we certainly cannot put their self-esteem on a bathroom scale. These kinds of variables are called **constructs** (pronounced CON-structs) and include personality traits (e.g., extroversion), emotional states (e.g., fear), attitudes (e.g., toward taxes), and abilities (e.g., athleticism).

Constructs cannot be observed directly. One reason is that they often represent tendencies to think, feel, or act in certain ways. For example, to say that a particular college student is highly extroverted (see Note 5.6 "The Big Five") does not necessarily mean that she is behaving in an extroverted way right now. In fact, she might be sitting quietly by herself, reading a book. Instead, it means that she has a general tendency to behave in extroverted ways (talking, laughing, etc.) across a variety of situations. Another reason constructs cannot be observed directly is that they often involve internal processes. Fear, for example, involves the activation of certain central and peripheral nervous system structures, along with certain kinds of thoughts, feelings, and behaviors—none of which is necessarily obvious to an outside observer. Notice also that neither extroversion nor fear "reduces to" any particular thought, feeling, act, or physiological structure or process. Instead, each is a kind of summary of a complex set of behaviors and internal processes.

Aspects of the Human Personality: The Big Five

The Big Five is a conceptual version of five broad dimensions that capture much of the variation in human personality. Each of the Big Five can even be defined in terms of six more specific constructs called "facets" (Costa & McCrae, 1992).[1]

Figure 5.1
The Big Five Personality Dimensions

Big Five Dimenson	Facets					
Openness to Experience	Fantasy	Aesthetics	Feelings	Actions	Ideas	Values
Conscientiousness	Competence	Order	Dutifulness	Achievement Striving	Self-Discipline	Deliberation
Extraversion	Warmth	Gregariousness	Assertiveness	Activity	Excitement Seeking	Positive Emotions
Agreeableness	Trust	Straight-forwardness	Altruism	Compliance	Modesty	Tender-Mindedness
Neuroticism	Worry	Anger	Discourage-ment	Self-Consciousness	Impulsivity	Vulnerability

A conceptual definition is an idea, or the basic concept that defines some phenomenon, but it is amorphous. Not yet well defined. Such things as depression, and anxiety, or concepts. But, concepts are generally complex and multifaceted.

The **conceptual definition** of a psychological construct describes the behaviors and internal processes that make up that construct, along with how it relates to other variables. For example, a conceptual definition of neuroticism (another one of the Big Five) would be that it is people's tendency to experience negative emotions such as anxiety, anger, and sadness across a variety of situations. This definition might also include that it has a strong genetic component, remains fairly stable over time, and is positively correlated with the tendency to experience pain and other physical symptoms.

Students sometimes wonder why, when researchers want to understand a construct like self-esteem or neuroticism, they do not simply look it up in the dictionary. One reason is that many scientific constructs do not have counterparts in everyday language (e.g., working memory capacity). More important, researchers are in the business of developing definitions that are more detailed and precise—and that more accurately describe the way the world is—than the informal definitions in the dictionary. As we will see, they do this by proposing conceptual definitions, testing them empirically, and revising them as necessary. Sometimes

they throw them out altogether. This is why the research literature often includes different conceptual definitions of the same construct. In some cases, an older conceptual definition has been replaced by a newer one that works better. In others, researchers are still in the process of deciding which of various conceptual definitions is the best.

Operational Definitions

In operationalizing we define the edges of the concept. It is an attempt to be precise, but often this precession is imposed. For example research in psychopharmacology, neuroscience, and psychotherapy are now strongly suggesting that there is more than one type of depression. So there is a certain amount of fine tuning or operational definitions as we learn more.

An **operational definition** is an attempt to me more precise with what the variable is in such a way that it can be measured. These measures generally fall into one of three broad categories.

1. **Self-report measures** are those in which participants report on their own thoughts, feelings, and actions, as with the Rosenberg Self-Esteem Scale.
2. **Behavioral measures** are those in which some other aspect of participants' behavior is observed and recorded. This is an extremely broad category that includes the observation of people's behavior both in highly structured laboratory tasks and in more natural settings. A good example of the former would be measuring working memory capacity using the backward digit span task. A good example of the latter is a famous operational definition of physical aggression from researcher Albert Bandura and his colleagues (Bandura, Ross, & Ross, 1961). [2] They let each of several children play for 20 minutes in a room that contained a clown-shaped punching bag called a Bobo doll. They filmed each child and counted the number of acts of physical aggression he or she committed. These included hitting the doll with a mallet, punching it, and kicking it. Their operational definition, then, was the number of these specifically defined acts that the child committed in the 20-minute period.
3. **Physiological measures** are those that involve recording any of a wide variety of physiological processes, including heart rate and blood pressure, galvanic skin response, hormone levels, and electrical activity and blood flow in the brain.

For any given variable or construct, there will be multiple operational definitions. Stress is a good example. A rough conceptual definition is that stress is an adaptive response to a perceived danger or threat that involves physiological, cognitive, affective, and behavioral components. But researchers have operationally defined it in several ways. The Social Readjustment Rating Scale is a self-report questionnaire on which people identify stressful events that they have experienced in the past year and assigns points for each one depending on its severity. For example, a man who has been divorced (73 points), changed jobs (36 points), and had a change in sleeping habits (16 points) in the past year would have a total score of 125. The Daily Hassles and Uplifts Scale is similar but focuses on everyday stressors like misplacing things and being concerned about one's weight. The Perceived Stress Scale is another self-report measure that focuses on people's feelings of stress (e.g., "How often have you felt nervous and stressed?"). Researchers have also operationally defined stress in terms of several physiological variables including blood pressure and levels of the stress hormone cortisol.

Converging Operations

When social scientist use multiple operational definitions of the same construct—either within a study or across studies—they are using **converging operations**. The idea is that the various operational definitions are "converging" on the same construct. When scores based on several different operational definitions are closely related to each other and produce similar patterns of results, this constitutes good evidence that the construct is being measured effectively and that it is useful. The various measures of stress, for example, are all correlated with each other and have all been shown to be correlated with other variables such as

immune system functioning (also measured in a variety of ways) (Segerstrom & Miller, 2004). [3] This is what allows researchers eventually to draw useful general conclusions, such as "stress is negatively correlated with immune system functioning," as opposed to more specific and less useful ones, such as "people's scores on the Perceived Stress Scale are negatively correlated with their white blood counts."

Levels of Measurement

The psychologist S. S. Stevens suggested that scores can be assigned to individuals so that they communicate more or less quantitative information about the variable of interest (Stevens, 1946). [4] For example, the officials at a 100-m race could simply rank order the runners as they crossed the finish line (first, second, etc.), or they could time each runner to the nearest tenth of a second using a stopwatch (11.5 s, 12.1 s, etc.). In either case, they would be measuring the runners' times by systematically assigning scores to represent those times. But while the rank ordering procedure communicates the fact that the second-place runner took longer to finish than the first-place finisher, the stopwatch procedure also communicates how much longer the second-place finisher took. Stevens actually suggested four different **levels of measurement** (which he called "scales of measurement") that correspond to four different levels of quantitative information that can be communicated by a set of scores.

Nominal Level

The **nominal level** of measurement is used for categorical variables and involves assigning scores that are category labels. Category labels communicate whether any two individuals are the same or different in terms of the variable being measured. For example, if you look at your research participants as they enter the room, decide whether each one is male or female, and type this information into a spreadsheet, you are engaged in nominal-level measurement. Or if you ask your participants to indicate which of several ethnicities they identify themselves with, you are again engaged in nominal-level measurement.

Ordinal Level

The remaining three levels of measurement are used for quantitative variables.

The **ordinal level** of measurement involves assigning scores so that they represent the rank order of the individuals. Ranks communicate not only whether any two individuals are the same or different in terms of the variable being measured but also whether one individual is higher or lower on that variable.

Interval Level

The **interval level** of measurement involves assigning scores so that they represent the precise magnitude of the difference between individuals, but a score of zero does not actually represent the complete absence of the characteristic. A classic example is the measurement of heat using the Celsius or Fahrenheit scale. The difference between temperatures of 20°C and 25°C is precisely 5°, but a temperature of 0°C does not considered to be measured at the interval level does not mean that there is a complete absence of heat.

Ratio Level

The **ratio level** of measurement involves assigning scores in such a way that there is a true zero point that represents the complete absence of the quantity. Height measured in meters and weight measured in kilograms are good examples. So are counts of discrete objects or events such as the number of siblings one has or the number of questions a student answers correctly on an exam. Below is a diagram showing the levels, which figuratively form an ascending staircase, from the most basic nominal level to the highest ratio level. One commonly used variable at the ratio level is money.