# Chapter 2: Psychological Measurement

Researchers Tara MacDonald and Alanna Martineau were interested in the effect of female university students’ moods on their intentions to have unprotected sexual intercourse (MacDonald & Martineau, 2002). 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.

To students who are just getting started in psychological research, the challenge of measuring such variables might seem insurmountable. Is it really possible to measure things as intangible as self-esteem, mood, or an intention to do something? The answer is a resounding yes, and in this chapter, we look closely at the nature of the variables that psychologists study and how they can be measured. We also look at some practical issues in psychological measurement.

## Do You Feel You Are a Person of Worth?

The Rosenberg Self-Esteem Scale (Rosenberg, 1989) 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 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.

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

The point of scales like this is to take an abstract construct and derive a quantitative measure that captures the idea of self-esteem and produce a number which is higher for participants who have greater self-esteem.

# Understanding Psychological Measurement

Learning Objectives

1. Define measurement and give several examples of measurement in psychology.
2. Explain what a psychological construct is and give several examples.
3. Distinguish conceptual from operational definitions, give examples of each, and create simple operational definitions.
4. Distinguish the four levels of measurement, give examples of each, and explain why this distinction is important.

## What Is Measurement?

Measurement is 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 using a meat thermometer. It is also consistent with measurement in the other 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/s2). The result of this procedure is a score that represents the object’s potential energy.

This general definition of measurement is consistent with measurement in psychology too. Psychological measurement is often referred to as psychometrics. Imagine, for example, that a cognitive psychologist wants to measure a person’s working memory capacity—their 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, in which she reads a list of two digits to the person and asks them 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 their 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 they have 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 the person’s current level of depression.

The important point here is that measurement does not require any particular instruments or procedures. What it does require is some systematic procedure for assigning scores to individuals or objects so that those scores represent the characteristic of interest.

# Psychological Constructs

Many variables studied by psychologists are straightforward and simple to measure. These include age, height, weight, and birth order. 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., extraversion), emotional states (e.g., fear), attitudes (e.g., toward taxes), and abilities (e.g., athleticism).

Psychological 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 university student is highly extraverted does not necessarily mean that she is behaving in an extraverted 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 extraverted ways (e.g., being outgoing, enjoying social interactions) across a variety of situations. Another reason psychological 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 extraversion 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.

## Conceptually Defining the Construct

Having a clear and complete conceptual definition of a construct is a prerequisite for good measurement. For one thing, it allows you to make sound decisions about exactly how to measure the construct. If you had only a vague idea that you wanted to measure people’s “memory,” for example, you would have no way to choose whether you should have them remember a list of vocabulary words, a set of photographs, a newly learned skill, an experience from long ago, or have them remember to perform a task at a later time. Because psychologists now conceptualize memory as a set of semi-independent systems, you would have to be more precise about what you mean by “memory.” If you are interested in long-term episodic memory (memory for previous experiences), then having participants remember a list of words that they learned last week would make sense, but having them try to remember to execute a task in the future would not. In general, there is no substitute for reading the research literature on a construct and paying close attention to how others have defined it.

## Example: Personality and The Big Five

The Big Five is a set 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): Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism.

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 is hypothesized to have a strong genetic component, is required to remain fairly stable over time, and is positively correlated with other types of measurements, such as 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 fits and works better. In others, researchers are still in the process of deciding which of various conceptual definitions is the best.

# Operational Definitions

Once you have a conceptual definition of the construct you are interested in studying it is time to operationally define the construct. Recall an operational definition is a definition of the variable in terms of precisely how it is to be measured. Since most variables are relatively abstract concepts that cannot be directly observed (e.g., stress), and observation is at the heart of the scientific method, conceptual definitions must be transformed into something that can be directly observed and measured. Most variables can be operationally defined in many different ways. For example, stress can be operationally defined as people’s scores on a stress scale such as the Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983), cortisol concentrations in their saliva, or the number of stressful life events they have recently experienced. As described below, operationally defining your variable(s) of interest may involve using an existing measure or creating your own measure.

An operational definition is a definition of a variable in terms of precisely how it is to be measured. These measures generally fall into one of three broad categories. Self-report measures are those in which participants report on their own thoughts, feelings, and actions, as with the Rosenberg Self-Esteem Scale (Rosenberg, 1965). 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). 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 the child 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 during the 20-minute period. Finally, 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 (Holmes & Rahe, 1967) 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 Hassles and Uplifts Scale (Delongis, Coyne, Dakof, Folkman & Lazarus, 1982) is similar but focuses on everyday stressors like misplacing things and being concerned about one’s weight. The Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983) 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.

When psychologists 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” or coming together 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). 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.”

## Experiment 1

For example, in the in-class experiment, we measured ‘memory’ by score on a recognition test where you saw a list of words and for each one responded whether you had seen it before. This produces a numerical measure of memory in the number of correct answers. However, it is fair to also say that there are lot of other ways to think about memory. Memory can refer to being able to recount the events of an experience you had yesterday. Another common way to measure memory is via tests of recall, e.g., asking participants to report all the words they had seen during the original study phase. This would also produce a quantitative measure of memory for the word list. In more advanced memory research, there are theoretical questions about how recognition and recall memory may be influenced by underlying mechanisms that might be specific to those processes. Recalling words seems to depend on something like “searching” our memories that might not be part of the process of deciding if you recognize a word seen before.

It would also be fair to say that any measure of memory for a list of arbitrary, unrelated words fails to capture important ideas that people are interested in that relate to the concept of “memory.” One of the most common complains about memory is memory failures, such as the challenging issue of remembering somebody’s name after you meet them. People will also have the experience of walking into a room and forgetting why you went into the room, which is also described as a failure of memory. Understanding factors that affect memory for lists of words may inform our understanding of these kinds of memory failures, but the distance from the operational definition employed in our experiment to those applications should be noted in considering the meaning of our findings.

All forms of science employ measurement, but the idea of the distance from the operational definition to the underlying concept is somewhat unique to psychological science. In other areas like biology, chemistry, or physics it is more commonly the case that there is less debate about what is being measured exactly. Because psychology is the science of people, we have the advantage of intuition and a basic understanding of the high-level concepts. We all know what words like ‘memory’ or ‘anxiety’ mean. However, when we design experiments or read about others’ experimental work, we need to identify more precise definitions that turns these conceptual ideas into numbers. This also highlights the complexity of a word like “memory” and the associated challenge of indicating exactly what aspect of memory is being incorporated into the operational definition. This complexity is also why much modern psychological research uses increasingly specific and precise terminology to capture sub-areas of interest. For example, if you are interested in research aimed at understanding the phenomenon of forgetting why you walked into a room, you will want to look for research on “prospective memory,” which is built around operational definitions based on memory for intentions to carry out actions and when that process surprisingly fails.

The process of establishing operational definitions applies to the process of setting up both the independent and dependent variables for a study. Many of the terms used to describe the key ideas in “measurement” apply more obviously to the dependent variable. For our basic experimental design, we expect the dependent variable to be a **measured operational definition**, which is a quantitative number that changes in a direction that can be conceptually connected to the construct. For our Experiment 1, more words recognized is clearly associated with more memory. It is also fine to consider measures that move the other direction, such as a measure like reaction time (speed to make a response) which tends to go down as a reflection of more knowledge. In communication about research, it is necessary to be clear about the details of the type and direction used for measurement.

## Levels of Measurement

The psychologist S. S. Stevens suggested that scores can be assigned to individuals in a way that communicates more or less quantitative information about the variable of interest (Stevens, 1946). 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 types of information that can be communicated by a set of scores, and the statistical procedures that can be used with the information.

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 ask your participants about their marital status, 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. The essential point about nominal scales is that they do not imply any ordering among the responses. For example, when classifying people according to their favorite color, there is no sense in which green is placed “ahead of” blue. Responses are merely categorized. Nominal scales thus embody the lowest level of measurement.

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. For example, a researcher wishing to measure consumers’ satisfaction with their microwave ovens might ask them to specify their feelings as either “very dissatisfied,” “somewhat dissatisfied,” “somewhat satisfied,” or “very satisfied.” The items in this scale are ordered, ranging from least to most satisfied. This is what distinguishes ordinal from nominal scales. Unlike nominal scales, ordinal scales allow comparisons of the degree to which two individuals rate the variable. For example, our satisfaction ordering makes it meaningful to assert that one person is more satisfied than another with their microwave ovens. Such an assertion reflects the first person’s use of a verbal label that comes later in the list than the label chosen by the second person.

On the other hand, ordinal scales fail to capture important information that will be present in the other levels of measurement we examine. In particular, the difference between two levels of an ordinal scale cannot be assumed to be the same as the difference between two other levels (just like you cannot assume that the gap between the runners in first and second place is equal to the gap between the runners in second and third place). In our satisfaction scale, for example, the difference between the responses “very dissatisfied” and “somewhat dissatisfied” is probably not equivalent to the difference between “somewhat dissatisfied” and “somewhat satisfied.” Nothing in our measurement procedure allows us to determine whether the two differences reflect the same difference in psychological satisfaction. Statisticians express this point by saying that the differences between adjacent scale values do not necessarily represent equal intervals on the underlying scale giving rise to the measurements. (In our case, the underlying scale is the true feeling of satisfaction, which we are trying to measure.)

The interval level of measurement involves assigning scores using numerical scales in which intervals have the same interpretation throughout. As an example, consider either the Fahrenheit or Celsius temperature scales. The difference between 30 degrees and 40 degrees represents the same temperature difference as the difference between 80 degrees and 90 degrees. This is because each 10-degree interval has the same physical meaning (in terms of the kinetic energy of molecules).

Interval scales are not perfect, however. In particular, they do not have a true zero point even if one of the scaled values happens to carry the name “zero.” The Fahrenheit scale illustrates the issue. Zero degrees Fahrenheit does not represent the complete absence of temperature (the absence of any molecular kinetic energy). In reality, the label “zero” is applied to its temperature for quite accidental reasons connected to the history of temperature measurement. Since an interval scale has no true zero point, it does not make sense to compute ratios of temperatures. For example, there is no sense in which the ratio of 40 to 20 degrees Fahrenheit is the same as the ratio of 100 to 50 degrees; no interesting physical property is preserved across the two ratios. After all, if the “zero” label were applied at the temperature that Fahrenheit happens to label as 10 degrees, the two ratios would instead be 30 to 10 and 90 to 40, no longer the same! For this reason, it does not make sense to say that 80 degrees is “twice as hot” as 40 degrees. Such a claim would depend on an arbitrary decision about where to “start” the temperature scale, namely, what temperature to call zero (whereas the claim is intended to make a more fundamental assertion about the underlying physical reality).

In psychology, the intelligence quotient (IQ) is often considered to be measured at the interval level. While it is technically possible to receive a score of 0 on an IQ test, such a score would not indicate the complete absence of IQ. Moreover, a person with an IQ score of 140 does not have twice the IQ of a person with a score of 70. However, the difference between IQ scores of 80 and 100 is the same as the difference between IQ scores of 120 and 140.

Finally, 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. You can think of a ratio scale as the three earlier scales rolled up in one. Like a nominal scale, it provides a name or category for each object (the numbers serve as labels). Like an ordinal scale, the objects are ordered (in terms of the ordering of the numbers). Like an interval scale, the same difference at two places on the scale has the same meaning. However, in addition, the same ratio at two places on the scale also carries the same meaning (see Table 4.1).

The Fahrenheit scale for temperature has an arbitrary zero point and is therefore not a ratio scale. However, zero on the Kelvin scale is absolute zero. This makes the Kelvin scale a ratio scale. For example, if one temperature is twice as high as another as measured on the Kelvin scale, then it has twice the kinetic energy of the other temperature.

Another example of a ratio scale is the amount of money you have in your pocket right now (25 cents, 50 cents, etc.). Money is measured on a ratio scale because, in addition to having the properties of an interval scale, it has a true zero point: if you have zero money, this actually implies the absence of money. Since money has a true zero point, it makes sense to say that someone with 50 cents has twice as much money as someone with 25 cents.

Stevens’s levels of measurement are important for at least two reasons. First, they emphasize the generality of the concept of measurement. Although people do not normally think of categorizing or ranking individuals as measurement, in fact, they are as long as they are done so that they represent some characteristic of the individuals. Second, the levels of measurement can serve as a rough guide to the statistical procedures that can be used with the data and the conclusions that can be drawn from them. With nominal-level measurement, for example, the only available measure of central tendency is the mode. With ordinal-level measurement, the median or mode can be used as indicators of central tendency. Interval and ratio-level measurement are typically considered the most desirable because they permit for any indicators of central tendency to be computed (i.e., mean, median, or mode). Also, ratio-level measurement is the only level that allows meaningful statements about ratios of scores. Once again, one cannot say that someone with an IQ of 140 is twice as intelligent as someone with an IQ of 70 because IQ is measured at the interval level, but one can say that someone with six siblings has twice as many as someone with three because number of siblings is measured at the ratio level.

Table 4.1 Features of Levels of Measurements

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Level of Measurement | Category labels | Rank order | Equal intervals | True zero |
| NOMINAL | X |  |  |  |
| ORDINAL | X | X |  |  |
| INTERVAL | X | X | X |  |
| RATIO | X | X | X | X |

## Reliability and Validity of Operational Definitions

Developing a novel measure of a construct that consistently and accurately numerically captures a complex construct is a complex and time-consuming task. We will discuss the general methodology for this later (Chapter 17, Surveys and Instrument Design) since this process is more often engaged with as part of non-experimental research than experimental research and is also generally outside the scope of this introductory class on psychological science. However, drawing inferences about experimental data will require considering how well the operational definition captures the underlying construct. Misalignment between the operational definition and the construct can lead to problems with inferences about the construct or can limit the applicability of findings to contexts outside the laboratory.

In the context of measurement, **reliability** refers to how consistently the measure obtains an accurate assessment of the underlying construct. For example, in personality research, characteristics such as ‘conscientiousness’ are expected to be stable individual traits over time. That means that subsequent attempts to measure the trait should generally produce the same number. However, data collected from human participants is virtually never perfectly stable for a wide variety of reasons. Participants might have external or internal distractions while engaged with a measure, or might have state-level effects (e.g., tiredness or hunger) that unexpectedly influence the score obtained. Everything that influences our measure that us unrelated to the construct creates **measurement error**, which shows up in our experimental data as a contribution to the observed variance in performance. We will discuss methodological techniques for managing measurement error as best we can in Chapters 3 and 4, but even with best practices, there will always be some component of “noise” in our data (also important for our statistical approach, Chapter 5).

Another key aspect of an effective measured operational definition is its **validity** in capturing the underlying construct. Robust techniques for establishing validity of a novel measure are complex (Chapter 17) but a simpler key version of the issue is seen as the **face validity** of a measure. Face validity is one that can often be evaluated intuitively and is simply a question of whether the measure actually relates to the underlying construct. If we were to claim that our Experiment 1 recognition memory measure is a measure of how likely you are to forget why you walked into the kitchen, we would lack face validity and this level of inference about our data should not be trusted. In contrast, if we claimed that our measure was relevant for understanding how students could build better memory for studying material in the classroom, we would have better face validity (but not perfect and examples of where there might be a disconnect is left as an exercise for the reader).

## Key Takeaways and Exercises

* Measurement is the assignment of scores to individuals so that the scores represent some characteristic of the individuals. Psychological measurement can be achieved in a wide variety of ways, including self-report, behavioral, and physiological measures.
* Psychological constructs such as intelligence, self-esteem, and depression are variables that are not directly observable because they represent behavioral tendencies or complex patterns of behavior and internal processes. An important goal of scientific research is to conceptually define psychological constructs in ways that accurately describe them.
* For any conceptual definition of a construct, there will be many different operational definitions or ways of measuring it. The use of multiple operational definitions, or converging operations, is a common strategy in psychological research.
* Variables can be measured at four different levels—nominal, ordinal, interval, and ratio—that communicate increasing amounts of quantitative information. The level of measurement affects the kinds of statistics you can use and conclusions you can draw from your data.
* Psychological researchers do not simply assume that their measures work. Instead, they conduct research to show that they work. If they cannot show that they work, they stop using them.
* There are two distinct criteria by which researchers evaluate their measures: reliability and validity. Reliability is consistency across time (test-retest reliability), across items (internal consistency), and across researchers (interrater reliability). Validity is the extent to which the scores actually represent the variable they are intended to.
* Good measurement begins with a clear conceptual definition of the construct to be measured. This is accomplished both by clear and detailed thinking and by a review of the research literature.

## Exercises

* Practice: Complete the [Rosenberg Self-Esteem Scale](https://www.wwnorton.com/college/psych/psychsci/media/rosenberg.htm) and compute your overall score.
* Practice: Think of three operational definitions for sexual jealousy, decisiveness, and social anxiety. Consider the possibility of self-report, behavioral, and physiological measures. Be as precise as you can.
* Practice: For each of the following variables, decide which level of measurement is being used.
  + A university instructor measures the time it takes her students to finish an exam by looking through the stack of exams at the end. She assigns the one on the bottom a score of 1, the one on top of that a 2, and so on.
  + A researcher accesses her participants’ medical records and counts the number of times they have seen a doctor in the past year.
  + Participants in a research study are asked whether they are right-handed or left-handed.
* Discussion: Think back to the last college exam you took and think of the exam as a psychological measure. What construct do you think it was intended to measure? Comment on its face and content validity. What data could you collect to assess its reliability and criterion validity?
* Practice: Write your own conceptual definition of self-confidence, irritability, and athleticism.
* Practice: Choose a construct (sexual jealousy, self-confidence, etc.) and find two measures of that construct in the research literature. If you were conducting your own study, which one (if either) would you use and why?

## References

Amir, N., Freshman, M., & Foa, E. (2002). Enhanced Stroop interference for threat in social phobia, 1–9.

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models, 575–582.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress.386-396.

Costa, P. T., Jr., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory, 5–13.

Delongis, A., Coyne, J. C., Dakof, G., Folkman, S., & Lazarus, R. S. (1982). Relationships of daily hassles, uplifts, and major life events to health status. (2), 119-136.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big Five personality domains, 504–528.

Holmes, T. H., & Rahe, R. H. (1967). The Social Readjustment Rating Scale. (2), 213-218.

Levels of Measurement. Retrieved from [http://wikieducator.org/ Introduction\_to\_Research\_Methods\_In\_Psychology/ Theories\_and\_Measurement/Levels\_of\_Measurement](http://wikieducator.org/%20Introduction_to_Research_Methods_In_Psychology/%20Theories_and_Measurement/Levels_of_Measurement)

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?, 299–306.

Rosenberg, M. (1989). (rev. ed.). Middletown, CT: Wesleyan University Press.

Segerstrom, S. E., & Miller, G. E. (2004). Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry, 601–630.

Stevens, S. S. (1946). On the theory of scales of measurement, 677–680.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions, 643–662.