Pre-lecture Notes I.4 – Measuring the Unobservable

Think about all of the issues that psychologists are interested in – these include depression, anxiety, attention, memory, problem-solving abilities, interpersonal attribution, etc. Notice how many of these cannot be directly observed; they are characteristics of the mind, not behavior, and, as of now, we don't have a way of reading people's minds. Given that psychology is an empirical science and, therefore, relies upon objective and replicable data, this is a problem.

One option, of course, is to disallow any discussion of anything that cannot be directly observed. Under this approach, you would never talk about something like depression, since it cannot be observed. Instead, you would only talk about the behavioral manifestations (or symptoms) of depression, because these can be observed. This is the approach taken by behaviorists and some neo-behaviorists.

The other approach, which is taken by a large majority of psychologists, is to create ways to convert or translate unobservable constructs, such as depression, into observable behaviors, such as responses on a questionnaire. In some cases, this translation is almost too obvious to discuss, such as defining the amount of time that it takes a person to mentally solve a problem as the lag or delay between the moment when they were given the problem and the moment when they give the (correct) answer. In other cases, however, which includes the definition of depression as the condensed score on a specific questionnaire, quite a bit of discussion and background research is needed.

This brings us to the first of the four types of validity that determine the value or quality of psychological research. Before getting into the details, however, a few things about the four types of validity should be said in advance. First, in each case, it will not be an all-or-none question of whether the given study, experiment, procedure, or method does or doesn't have a certain type of validity; it will always be a question of how much. All-or-none statements about validity are too simple to ever be true. A given study, experiment, procedure, or method has a certain level of each type of validity that is rarely zero or 100%. For this reason, all of the definitions for the four types of validity will start with "the extent to which..." as opposed to "whether or not...."

Second, while each of the four types of validity is distinct, they are by no means independent of each other. In fact, in several cases, the members of a pair of validities will be in direct competition with each other such that more of one type of validity will automatically imply less of another. Therefore, researchers are often faced with an unpleasant decision: which type of validity will be maximized and which type of validity will be allowed to (or forced to) go down. More generally, the fact that some pairs of validities are in direct competition leads to the following statement that applies to all forms of psychological research: *there is no such thing as a perfect experiment*. This is true because it is not possible to have maximum levels of all four types of validity at the same time.

Turning now to the measures that psychologists use to estimate hidden, unobservable, mental constructs, such as depression, it all comes down to something called an "operational definition." Here's your first to-be-memorized definition:

Operational Definition -a statement that maps one or more empirical measures onto one or more theoretical constructs

For example, the standard operational definition of depression is "the score that the subject receives on the Beck Depression Inventory (BDI)" (which is a 21-item questionnaire first published in the early 1960s and revised several times since). To be clear, when using this operational definition, the score on the BDI *is* the subject's depression. When you take an operational definition super seriously, you don't say that the subject has a certain level of depression and the BDI provides a good estimate of it, because that way of saying it still includes something that is unobservable; instead, you actually talk about the BDI score as being the subject's depression. In this way, what you are interested in has ceased to be something unobservable and has become something objective and replicable. It doesn't require any secret skill to use the BDI questionnaire, so it's objective. And anyone can collect the same data using the BDI, so it's replicable.

With all that said, because most psychologists are a tad nervous about assuming that their operational definitions are perfect, we do often say that measures provide estimates of theoretical constructs. But please don't take this as implying that psychologists are being slimy or hedging their bets; take it in a positive way: we are aware that our current operational definitions are the best we've developed so far and we are open to suggestions as to how to make them better.

[deep breath]

With regard to the first type of validity – which is called "construct validity" – a cheap and easy way to define this concept is: *the extent to which the operational definition being used is accurate*. But please don't use this definition as it doesn't really help since it depends too much on the definition of "operational definition" (which most people won't know). Instead, at a minimum, use this:

Construct Validity [OK definition] – the extent to which the measure provides an accurate estimate of the target theoretical construct

Note how construct validity concerns the relationship between some measure (e.g., BDI score) and some theoretical construct (e.g., depression); this makes it very much like an operational definition, which is why some people are tempted to use the cheap and easy definition of construct validity. But, again, please don't do this because it's much too simple. Most of all, using either the cheap and easy definition of construct validity or even the slightly better one above leaves out a very, very important point: that construct validity has two different components.

A much better definition of Construct Validity is a twist on the oath that people must swear before testifying in court: "do you swear to tell the truth, the whole truth, and nothing but the truth?":

Construct Validity [best definition] – the extent to which the measure provides an exhaustive and selective estimate of the target theoretical construct

In the best definition of construct validity, the word "exhaustive" is used to say that the measure should cover *all* aspects of the target construct; nothing is left out; it's the whole truth. Meanwhile, the word "selective" here says that the measure *only* covers things that are part of the target construct; nothing that isn't part of the target construct is included by mistake; it's nothing but the truth.

In lecture, I will trace the history of construct validity, showing you how we got to the final definition.