Measuring Psychological Constructs ~ Operational Definitions

**Measurement** the systematic assignment of numbers to objects or attributes of objects. In other sciences, measurement is generally not as difficult a problem as it is in psychological research. In chemistry and biology, for example, there are clear-cut definitions of the variables of interest. For example, growth is measured in millimeters or kilograms, color can be defined by wavelength, and velocity is defined as the rate of change of position of an object with respect to time. In psychology, we are not as lucky. Most of the **Constructs** we measure are **abstract**. We see evidence of them in human behavior, but we cannot measure them directly.

E.g., Do Intelligence Tests measure “intelligence”?

Once you have picked a research question to address, the next problem is deciding how to measure your **constructs**. The constructs in your study are the aspects of human behavior or quality that you want to study. Research requires a means of quantifying the constructs. The manner in which you measure each construct is the **operational definition** of the construct for your study. For example, you might be interested in studying the relationship between people’s self-confidence and their plans to continue onto graduate school. How would you measure these constructs? Is there one and only one way to do it?

E.g., **Does TV violence increase aggression in children?**

**Operational Definition:** defining the variables according to how they are measured.

- Give precise criteria that clearly communicate what you are measuring.
- The usefulness and interpretability of your study depends largely on how meaningfully you are able to measure your variables.
- Your literature review will help you to consider your options.

For many variables of interest to psychologists, there is no “one”, “perfect” or “totally accepted” measure. This means that it is very important to pay attention to the operational definition of your variables. You will need to defend your choice.

- keep in mind your population.
- keep in mind the limitations of the project you are doing.
- keep ethical concerns in mind.

**When making a decision about how to measure your constructs, the validity and reliability of the measurement must be considered.** A test is **valid** when it measures what it is required to for the purposes of the study. A test is **reliable** if it yields consistent results.
A test can be both reliable and valid, one or the other, or neither. Reliability is a prerequisite for measurement validity.

“Everything boils down to validity.” —Pelham & Blanton, 2003

**Construct Validity**: The degree to which the operationalization of the variables accurately measures what they are intended to measure.

**Types of Construct Validity**

- **Face Validity** – Does it appear that the measurement assesses the construct under study? Some measures have very low face validity – yet are very high in other types of construct validity.

- **Content Validity**: Is the full content of a concept’s definition included in the measure? Does it include a broad sample of aspects of the construct tested?

  To assess this you need to have a very clear idea of the concept you are trying to capture with the measure. For example, if you are interested in measuring risk taking what kinds of risk taking are you interested in (legal vs. illegal; instrumental vs recreational). There are many ways a person can be a risk taker. Does your measure cover all the ways that are relevant for your study?

- **Criterion Validity** (also called convergent validity): Is the measure consistent with other measures of this or related constructs.

  Two subcategories: predictive and concurrent

  **Predictive validity**: Predicts a known association between the construct you are measuring and something a future event. (i.e., does a measure of aggression in youth predict number of violent penalty minutes for hockey players).

  **Concurrent validity**: Associated with pre-existing indicators; something that already measures the same concept.

- **Discriminant Validity**: Does the measurement guard against contamination by unrelated constructs. For example, many self-report questionnaires are subject to social desirability effects. People tend to answer in ways that they feel are socially acceptable rather than providing honest answers about their behaviors, attitudes and beliefs. There are ways to measure social desirability that can be included in self-report inventories.
Sometimes we observe **behavioral, or performance outcomes**. Other times, we use **self-report or peer-report inventories**. In any case, when you research the background literature for your study, pay careful attention to how the construct is operationally defined by others. Often research publications will provide information on validity.

**Types of Reliability**

**Inter-observer**: There are consistent results among testers or coders who are rating the same information. Measuring agreement using the Inter-observer reliability coefficient is a good rule of thumb. A coefficient greater .80 indicates that the ratings have acceptable inter-observer reliability.

**Test-retest**: If we use the measure at two different times with no treatment in between, obtaining consistent results indicated test-re-test reliability. A coefficient of the relationship between scores at each testing should be great than .80 to indicate acceptable test-re-test reliability.

**Parallel-forms**: Two tests of different forms that supposedly test the same material will give the same results.

**Split-half reliability**: If the items are divided in half (e.g., odd vs. even questions) the two halves give the same results.

**Appropriateness of the Measure for your Population**:

**Sensitivity**: the degree to which your measure is able to define useful and valid distinctions between conditions. If you are looking for differences between (or relationships between) conditions, you may fail to find any because the measurement you use are not **sensitive** to the change that does occur.

Things to consider:

- Can your participants actually carry out the measurement task?
  - Can they understand the words used?

- If the task is at too low a level, your participants may not take it seriously.

- **Scale-attenuation effects** - difficulties in either interpreting results when performance on the variable is nearly perfect (a ceiling effect) or nearly lacking altogether (a floor effect). There will be no variation in your measures. Almost everyone scores at the same level. **Ceiling effects** result when the measure is so easy that even if a person gets better at the skill involved, their score will not change, because they are already scoring as high as the scale is capable of measuring. **Floor effects** result when the measure is so difficult that even if a
person gets better at the skill involved their score will not change, because even with improvement in the skill they still are unable to do the task well enough to increase their score.

Types of Measures

- **Discrete or Categorical Variables** - come in whole units or “categories”.
- **Continuous Variables** - form a “continuum” with fractional values possible.

Continuous scales can be converted to discrete scales. E.g., Height in inches to “Short”, “Average”, and “Tall”. Doing so however makes the measurement less precise.

- **Qualitative variables** - vary in “kind” or “type”
- **Quantitative variables** - vary in “amount”.

Types of Scales

Quantitative analysis is concerned with making sense of data, which are expressed in terms of quantities (numbers). Data comes in several forms. The type of descriptive statistics and analysis you will use for your study will depend on the nature of variables in your study. We will begin by discussing the characteristics of measurement scales.

Discrete (categorical) Variables

**Nominal Variables.** Numbers are used to label categories that differ in type. The magnitudes of the numbers are **not** meant to indicate that one category is better than another. No rank ordering of categories is intended. A common example of a nominal category is Sex. We might enter the number one to indicate that a subject is male and the number two to indicate that a subject is female. The number should only be considered a label (name) and should not be viewed as a value.

**Ordinal Variables.** Numbers in this case are meant to indicate an order to the levels of the variable being measured. For example, education level can be measured on an ordinal scale. We might use a zero to indicate that a subject has not completed a high school degree; a one to indicate that a person’s highest level of education is a high school degree; a two to indicate an associates degree; a three to indicate a B.A. or B.Sc.; a four to indicate an MA or M.Sc, and a five to indicate a Doctorate. In this case, the higher the number is the more education a person has. An Ordinal scale defines not only a difference between categories but also assigns an order to the variables. The intervals, however, on an Ordinal scale are not equal. By this I mean that you could not logically argue that a MA (labeled 4), is as much greater than a BA (labeled 3), as a Doctorate (labeled 5) is greater than an MA.
Continuous Variables

Interval Variables. Numbers on this scales indicate differences and order, however, they also have equal intervals. The magnitude of the categories indicates difference, order and degree of difference. An example of an interval scale is temperature measured in either Fahrenheit or Celsius. In this case, we can say that 50 degrees is as much higher than 25 degrees as 75 degrees is higher than 50 degrees. Thus, the intervals indicated by the numbers are equal. An interval scale, however, does not have an absolute zero. In our example, zero does not indicate an absence of temperature; therefore, we cannot say that 50 degrees is twice as hot as 25 degrees.

Ratio Scales. Numbers on this scale indicate differences, order, equal intervals and have an absolute or naturally falling zero point. For example, time measured in minutes is a ratio scale. Zero indicates no time. The importance of having a natural falling zero is that it allows us to make ratio comparisons (4 is twice as much as 2).