Validity Overview

In the context of research design, there are several types of validity but all are concerned with the *confidence you can have regarding certain conclusions*. You might consider the issues surrounding validity as the accumulated wisdom of countless researchers who learned the hard way how *not* to design a study. By understanding validity, you will not only be able to avoid making critical errors in your own research design, but also be able to evaluate how trustworthy the results from a particular study are. The three types of validity that we will discuss are **construct validity**, **internal validity**, and **external validity**.

Construct Validity

To understand construct validity, it is helpful to first review some vocabulary concerning variables. A **variable** is an abstract representation of a phenomenon (e.g., fear, intelligence, conformity) that can exist at two or more levels (it can *vary*). An **independent variable** is the variable that a researcher suspects is the cause of some behavior. For example, in an experiment on the effect of expertise of spokesperson on persuasion, expertise is the independent variable because as it takes on different levels (e.g., medical doctor, average person-on-the-street), the researcher expects that persuasion will change. In an experiment, the independent variable is **manipulated** (changed systematically) by the researcher. The different values that the independent variable takes on are often called *levels*, *conditions* or *groups*. For example, a researcher could manipulate the independent variable of expertise by assigning some participants to watch a commercial where the spokesperson is a medical doctor (the "doctor condition"), and others to watch the same commercial where the spokesperson is presented as an average person (the "average person" condition). A special case of an independent variable is a **subject variable**. A subject variable is a characteristic of the participants in your study, which usually cannot be manipulated by a researcher. Examples of subject variables are gender, personality traits such as extroversion, or grade point average. A **dependent variable** is the behavioral response from participants that is observed. For example, in an experiment on the effect of expertise on persuasion, a person's change in attitude (the degree to which they have been persuaded) is the dependent variable. To help you to identify the independent and dependent variables in a study, try to rephrase the study in terms of "The effect of *IV* on *DV*." Many studies can be phrased this way. Some people also remember the difference by remembering that the dependent variable *depends upon* (is influenced by) the independent variable.

Variables exist at both a conceptual level and a concrete or real-world level. For example, in a study on the effect of teacher expectations on student intelligence, the independent variable is teacher expectation and the dependent variable is student intelligence. These variables represent abstract concepts that cannot be directly measured or observed. In the study, researchers must find a way to bring the abstract concept down to earth and turn it into something that can be measured or manipulated. This concrete, real-world definition of a variable is called its **operational definition**. For example, intelligence may be operationally defined as a student's performance on the Wechsler Intelligence Scale for Children (WISC). Operational definitions are explicit statements of how abstract variables will be measured or manipulated in a study. They are sometimes the most challenging aspect of a study and often require a great deal of creativity and insight. For example, Harmon-Jones and Sigelman (2001) developed a clever measure of aggression. They allowed participants to choose what another person (actually an actor who had just insulted them) will receive in their water for a taste test: 1-3 tablespoons of sugar or vinegar. Operational definitions are an important part of the research process because they provide enough detail to permit other researchers either to critique how the study was done or to replicate it.

Construct validity is the confidence you can have that the operational definition faithfully represents the abstract construct it is supposed to. Sometimes, the whole purpose of a study is to demonstrate the construct validity of a new measure, for example of intelligence. Even when this is not the case, construct validity is critically important to any study because a study is only as good as its operational definitions. Researchers can have a very well-designed study with thousands of participants, but if they operationally define intelligence as head circumference (which is not correlated with intelligence), their results will be meaningless. There are four major ways of establishing construct validity, and each is its own type of validity: face validity, convergent validity, predictive validity, and discriminant validity.

Face validity. Face validity is a very rough and superficial assessment of an operational definition: does it *look like* it is measuring what it claims to measure? For example, if you pick up a test that claims to measure attitudes toward gun control and it has questions such as "Do you like cheese?", then you may consider the test to have poor face validity.

Convergent validity. Convergent validity is the degree to which your operational definition is correlated with variables that you would expect it to be correlated with. For example, let's say that a researcher wants to administer an intelligence test to people in another culture. To determine whether the translated test is still measuring intelligence, the researcher could correlate people's scores on the new test to outcomes that would be expected to be correlated with intelligence: years of education, reputation as an intelligent person, ability to solve practical problems quickly and accurately, grade point average in school, etc. If these variables are positively correlated with the new translated measure, you can have more faith that it is a legitimate measure of intelligence.

Predictive validity. A special case of convergent validity is predictive validity: the degree to which an operational definition predicts an outcome that it was designed to predict. For example, the Scholastic Aptitude Test (SAT) is designed to predict success in college. One measure of its predictive validity is the correlation between students' SAT scores and their grade point average in the first semester of college.

Discriminant validity. Discriminant validity is the degree to which the operational definition is able to discriminate between the target construct and closely related (but conceptually distinct) variables. Whereas convergent validity hopes for high positive correlations between the operational definition and related variables, discriminant validity hopes for correlations between operational definitions and distinct variables that are close to zero. For example, let's say that you come across a test of mathematical intelligence that includes word problems. You may be concerned that the test does not adequately discriminate between mathematical and verbal ability, that it is "contaminated" by verbal ability. If the correlation between this test of mathematical intelligence and a separate test of verbal ability is close to zero, then you can be much more confident that the math test has discriminant validity with regard to verbal intelligence. If the correlation between the math test and a verbal test is high and positive, then the math test is not adequately discriminating between mathematical and verbal ability.