# TOPIC 25 INTRODUCTION TO VALIDITY

The generic term for any type of measurement device (e.g., test, questionnaire, interview schedule, or personality scale) is **instrument**. Thus, for example, one might ask the question: "What type of *instrument* did the researcher use to collect the data?" In formal research reports, you will usually find the heading **instrumentation**, which is the name of the section of the report where the measurement devices used in the research are described. An important issue that is often addressed in this section is the question of **validity**.

We say an instrument is **valid** to the extent that it measures what it is designed to measure and accurately performs the function(s) it is purported to perform. For instance, consider an achievement test that emphasizes knowledge of facts on the westward movement in the United States. It will be only *modestly valid* when administered to students whose instruction emphasized critical appraisal of historical evidence on that period in history. Likewise, a typing test used by an employer is likely to be only *partially valid* if it is administered to applicants who are applying for a job that includes filing in addition to typing.

It is important to note that validity is *relative* to the purpose of testing. If the purpose is to measure achievement of students exposed to instruction on critical thinking, a test that measures only factual knowledge will be largely lacking in validity. For the purpose of measuring the achievement of students exposed to instruction in which the acquisition of factual knowledge was emphasized, the same test will be much more valid. Thus, before we can have an informed discussion of the validity of a particular instrument, the purpose for testing must be clearly stated.

Also, note that validity is a *matter of degree*. Therefore, it is appropriate to discuss *how valid* a test is—not *whether* it is valid. Given the imperfect state of measurement practice, it is safe to say that no test is perfectly valid.

Let us briefly explore some reasons why perfect validity eludes us. First, almost all tests tap only a *sample* of the behavior underlying the constructs we are trying to measure. Consider the construct of aptitude for college (i.e., having the abilities to succeed in college). College aptitude tests emphasize verbal and mathematical skills and leave untapped other skills that relate to success in college, such as ability to use a personal computer to do homework, knowledge of how to use an academic library, command of effective study skills, and having the maturity to persist in the face of difficulty in a course of instruction. Even within the domains that are tapped by a college aptitude test, only a small sample of verbal and mathematical problems can be presented within a test of reasonable length. Just as when we sample participants from a population (see Part C of this book), some samples are better than others, and all samples are subject to error.

Another reason why perfect validity eludes us is that some traits we want to measure are inherently elusive. Consider cheerfulness. We have all known people whose cheerfulness is contagious. However, even though we know it when we see it, how can we measure this trait in a systematic way in order to study it? We could, for example, ask a series of questions on how participants interact with others in various types of contexts, how they view adversity, etc. On the other hand, we might observe participants and rate them on the cheerfulness they exhibit in their interactions with others (e.g., Do they smile? Is the tone of their voice upbeat? and so on). While these procedures probably tap certain aspects of cheerfulness, they fail to capture the *full essence* of the trait. This illustrates the old principle that often the whole is greater than the sum of its parts. To compound the problem, we often examine only some of the parts when we measure a given trait (e.g., we look at only certain behaviors that indicate cheerfulness, and we do this in only certain public settings).

The problem of elusiveness, at first glance, seems to plague those with a quantitative orientation more than those with a qualitative orientation because quantitative researchers seek to reduce elusive constructs, such as cheerfulness, to numerical scores. Qualitative researchers, on the other hand, tend to measure in ways (such as unstructured interviews) that yield words to describe the extent to which traits are present. Yet unless qualitative researchers refer to specific behaviors and events in their reports, they will fail to describe results in enough detail so readers can picture the meanings that have been attached to a construct such as cheerfulness. Thus, listing events, quoting participants, and describing specific interactions, while less artificial than numerical scores, can also miss the essence of a trait, such as the feeling you get when you are with a genuinely cheerful person.