

Topic #2

RESEARCH VALIDITY

A key criterion in evaluating any test, measure, or piece of research is validity.

Define **VALIDITY** as—the APPROPRIATENESS of INFERENCES drawn from DATA

1. data or observations—not impressions or opinions
2. drawing inferences
3. appropriateness—implies purposes for which inferences are drawn
 - e.g., use of information posted at Facebook and other social media sites by employers to make hiring decisions (i.e., predictions about future job performance)?

The concept of validity is used in two different ways:

1. Research validity
 - (a) internal
 - (b) external
 - (c) statistical conclusion
 - (d) construct
2. Test and measurement validity
 - (a) criterion-related
 - (b) content-related
 - (c) construct-related

Research Validity—A conclusion based on a research study is valid when it corresponds to the actual or true state of the world.

Four facets or dimensions of research validity are commonly recognized—internal validity, external validity, statistical conclusion validity, and construct validity.

1. **Internal Validity**—is the extent to which we can infer that a relationship between two variables is causal or that the absence of a relationship implies absence of cause.
 - That is, is the system of research internally consistent? Do the relationships obtained follow from the research design?
 - In terms of experimental research designs, a study has internal validity if a cause-effect relationship actually exists between the independent and dependent variables.
 - The difficulty is determining whether the observed effect is caused only by the IV, since the DV could have been influenced by variables other than the IV.
 - **Extraneous Variable**—any variable other than the IV that influences the DV.
 - **Confounding**—when an extraneous variable systematically varies with variations or levels of the IV.
 - Internal validity is primarily determined by the quality of the research design, which is in turn determined by the degree or amount of experimental control (of extraneous and confounding variables).
 - **Threats**
 - A. **History** (events outside the lab)—the observed effects between the independent and dependent variable might be due to an event which takes place between the pretest and posttest when this event is not the treatment of research interest (e.g., effects of success/failure [IV] on feelings of depression [DV] with success condition run on a sunny day and failure on a gloomy, dark, cold, rainy day [weather = extraneous variable]).
 - B. **Maturation**—a source of error in a study related to the amount of time between measurements; concerned with naturally occurring changes in research participants (e.g., developmental or gerontological psychology research).
 - C. **Testing**—effects due to the number of times particular responses are measured—familiarity with the measuring instrument (e.g., increased scores on 2nd test).
 - D. **Attrition or Mortality**—the dropping out of some participants before a study is completed, causing a threat to validity (e.g., effect of learning strategies [IV] on end of semester grades [DV]; effect of attrition as a result of bad grades would be an extraneous variable).

- E. **Selection**—many studies compare two or more groups on some dependent variable after the introduction of an IV. Other studies like surveys just assess attitudes or opinions on an issue. In either case, sampling or selection into the study is critical. Samples must be comparable—in multi-group designs—or must represent the population (e.g., survey of attitudes towards endangered species—one would probably obtain very different results as a function of sampling from individuals in the logging industry vs. members of the Society for the Protection of Baby Seals).
- F. **Regression effects**—tendency of participants with extreme scores on first measure to score closer to the mean on a second testing; a statistical threat (e.g., scores on the 2nd test regress—move either higher or lower—to the true score).

- *These threats are corrected for by randomization*

2. **External Validity**—is the inference that presumed causal relationships can be generalized to and across alternate measures of cause and effect, and across different types, persons, settings, and times. That is, how generalizable are findings?

- The concern is whether the results of the research study can be generalized to another situation—specifically, participants, settings, and times.

- **Threats**

A. Other participants (interaction of selection and treatment)—**population validity**

Psychological research studies often use samples of convenience—"the experimentally accessible population". Consequently, the question is—How representative is the typical sample of the focal group of interest when participants are often chosen by availability?

The criticality or severity of convenient, nonrepresentative, nonrandom samples as a methodological flaw is to some extent a function of the topic domain being researched.

B. Other settings (interaction of setting and treatment)—**ecological validity**

C. Other times (interaction of history and treatment)—**temporal validity**

- *External validity may be increased by random sampling for representativeness*
- Importance of trade-off issues between internal and external validity.

3. **Statistical Conclusion Validity**—appropriateness of inferences (or conclusions) made from data as a result (or function) of conclusions drawn from statistical analysis. That is, are the IV and DV statistically related?

- **Threats**

- A. Low statistical power—this is the ability (power) of a statistical test to detect or identify relationships when they are actually present. All things being equal, the larger the sample size, the greater the power.

- B. Violated assumptions of statistical tests

- C. Reliability of measures' scores

- *These threats can be addressed by having adequate power, meeting the assumptions of tests, and using measures with acceptable levels of score reliability.*

4. **Construct Validity**—has to do with labels that can be placed on what is being observed and the extent to which said labels are theoretically relevant.

- Construct validity is a question of whether the research results support the theory underlying the research. That is, is there another theory that could adequately explain the same results?

- e.g., in polygraph research, is "ANXIETY" a better label than "LYING" for what is being studied?

- If the labels being used are irrelevant to the theory being researched, then the study can be said to lack construct validity.

- **Threats**

- A. Loose connection between theory and study.

- B. Changes in research participants' behaviors that result from their tendency to alter their behavior because they are being studied. These include but are not limited to effects such as:

- "good-subject" response
 - Hawthorne effect
 - social desirability responding [impression management]
 - evaluation apprehension
 - responses to experimenter expectancies

- *These effects can be controlled or minimized by using the following procedures:*

- (a) *Double-blind procedures*

- (b) *Single-blind procedures*

- (c) *Deception*

- Need to realize that the four facets or dimensions of research validity are interrelated and NOT independent of one another.

For example:

- statistical conclusion validity is necessary for demonstration of other types of validity
- internal validity must be achieved for construct validity to be obtained
- and external validity depends in part on the demonstration of at least statistical conclusion validity and internal validity

SUMMARY AND OVERVIEW OF SOME FREQUENTLY MISUNDERSTOOD AND CONFUSED TERMS AND CONCEPTS

- **Test.** Can or is used to refer to both a measurement tool or device (i.e., any means of operationalizing or quantifying variables [e.g., the GRE is a test/measure of scholastic aptitude or achievement]) OR a statistical significance test (e.g., the *t*-test as a statistical test for differences between means).
- **Study.** A test (i.e., measurement tool) is **not** a study. A study is an investigation or an assessment of the relationships between variables. A study is not a "test of the variables".
- **Qualities of a good test** (i.e., measurement tool or any means of operationalizing variables). Two properties commonly used to assess the quality of a test are the reliability of its scores, and (test and measurement and psychometric) validity (the appropriateness of the inferences drawn from the test's scores).
- **Qualities of a good study.** A good study must display relatively high levels of research validity (i.e., internal, external, statistical conclusion, and construct).
- Thus, in summary, **reliability and psychometric validity [which we will discuss next] are properties of test scores. However, research validity is a property of research studies.**