their work and endeavors to eliminate or at least minimize such distortion.

Maximize Precision

The third guideline involves accuracy in how something is measured. Although *precision* is a very general word, and takes on different meaning depending on what is being measured, validity is threatened in each step of the measurement process if investigators do not work with a great deal of care. It is easy, for example, to miss important indicants of a concept if an instrument is overly rigid in how it functions. Nevertheless, small inaccuracies in how an instrument is used can sometimes foster monumental levels of bias in the measurement process. In time and with experience, investigators learn how to assess where they can and cannot take liberties in how indicants are gathered, recorded, analyzed, and stored.

A Creative Endeavor

Strong validity arguments are made using formal logic. Such logic might not feel natural to individuals who prefer to change topics, research questions, or measurement procedures frequently. Nevertheless, researchers have acquired methods for overcoming such impulsiveness well enough to sustain high levels of creativity as they build defensible validity arguments.

Theresa A. Thorkildsen

See also Concurrent Validity; Construct Validity; Content Analysis; Content Validity; Criterion Validity; Critical Theory; Discriminant Analysis; Ecological Validity; Generalizability Theory; Psychometrics

Further Readings

Cronbach, J. L. (1988). Five perspectives on validity arguments. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3–18). Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, J. L., & Meehl, P. E. (1955). Construct validity in psychology tests. *Psychological Bulletin, 52,* 281–301.

Rorty, R. (1985). Solidarity or objectivity? In J. Rajchman & C. West (Eds.), *Post-analytic philosophy* (pp. 3–19). New York: Columbia University Press.

Thorkildsen, T. A. (2005). *Fundamentals of measurement in applied research.* Boston: Allyn & Bacon.

Validity of Research Conclusions

Sometimes described as "statistical conclusion validity," the validity of research conclusions refers to the degree to which the conclusions made about the null hypothesis are reasonable or correct. Because the null hypothesis typically states that a relationship between two variables does not exist, the validity of a research conclusion also refers to whether a relationship exists between two variables. Although the validity of research conclusions is distinct from construct validity and external validity, it is important to distinguish conclusion validity clearly from internal validity. Internal validity involves whether a relationship between two variables is a plausibly causal one. The validity of a research conclusion is concerned only with the presence or absence of a relationship between two variables. Thus, conclusion validity answers the most basic of questions from a cumulative set of validity questions (followed by questions regarding internal validity, construct validity, and then external validity), as it requires only a decision regarding covariation (not casual aspects such as temporal precedence of the presumed cause occurring prior to the presumed effect and minimal alternative explanations).

The validity of research conclusions is often considered an issue of statistical inference. For instance, a researcher who conducts a study of the effect of a persuasion technique on attitudes about undocumented immigration will want to conclude whether a relationship exists between the presences or the absences of the persuasive technique and change (or lack thereof) in the attitudes. However, the validity of research conclusions is also relevant for qualitative or observational field research. For instance, a researcher who observes traffic patterns and acts of vehicular aggression might want to conclude whether a relationship exists between drivers who slow down to observe nearby accidents and the likelihood of additional auto accidents. Despite the fact that the conclusions of a qualitative study might be based on impressionistic data, the validity of the research conclusions might be assessed, that is, whether a reasonable conclusion has been made about the relationship between two variables.

Possible Conclusions and Possible Consequences

To understand fully how it is that research conclusions might be considered valid, one must first understand the basic logic of hypothesis testing. Essentially, there are only two possible conclusions to all research endeavors. The first possible conclusion is that the data provide a reasonably significant result (in quantitative research, such as correlational and experimental studies, this is referred to as a *statistically significant* result; the term *reasonably significant* is used here to encompass both quantitative and qualitative research). That is, the data are sufficiently discrepant from the relationship stated by the null hypothesis that the null hypothesis is rejected. Because the null hypothesis typically asserts that a relationship between two variables does not exist, the significant result asserts that a relationship does exist.

The only other conclusion that might be made is that the data provide a nonsignificant result. That is, the data are not sufficiently discrepant from the relationship stated by the null hypothesis. Thus, the null hypothesis is retained. Note that the null hypothesis is not accepted. In other words, a nonsignificant result indicates uncertainty as to whether a relationship exists and does not necessarily indicate that no relationship exists.

There are only two possible consequences of the two possible conclusions, and each has two possible routes. First, the research conclusions might be correct. Retaining a null hypothesis that is "true" (i.e., one that cannot be rejected), or rejecting a null hypothesis that is false, are two ways of reaching a correct research conclusion. Second, the research conclusions might be incorrect. The two ways of reaching incorrect research conclusions are to either reject a null hypothesis that is "true" (Type I error) or retain a null hypothesis that is false (Type II error).

All quantitative research conclusions are made with respect to a particular level of statistically based confidence. By convention, researchers use the .05 level of statistical significance as a criterion for valid research conclusions (assuming that the design, implementation, and analysis of the research are free from the threats to validity described in the following discussion). This means that the probability that the conclusion about a significant result is false (Type I error) is less than 5%. On average, the probability that the conclusion about a nonsignificant result is false (Type II error) is approximately 20%. The discrepancy in the probabilities of Type I and Type II errors reflects the fact that the significance criterion and power are positively related and that researchers are more concerned about making Type I errors than they are about making Type II errors.

Threats to the Validity of Research Conclusions

Several statistical factors and research design features can affect the validity of research conclusions. Simply put, any factor that increases the likelihood of making either a Type I error or a Type II error will reduce the validity of the research conclusions. Although the probability of a Type I error is ultimately held at the discretion of the researcher, conventional use dictates that this probability be held at the .05 level (or less) of statistical significance.

Factors that decrease statistical power increase the likelihood of making a Type II error. Thus, these same factors can reduce the validity of research conclusions. These threats, which relate to power, include a small sample size, a more conservative criterion for determining statistical significance, large variance in the criterion variable (as a result of employing heterogeneous groups of subjects, failing to control for extraneous variables, or failing to employ reliable methods of manipulating and measuring variables), and employing the wrong statistical test (e.g., using a nondirectional test when a directional test is more appropriate).

Every statistical analysis depends on the nature of the variables manipulated and measured (e.g., categorical or continuous) and on several statistical assumptions. Thus, violation of the assumptions of a statistical test also serves as a threat to the validity of research conclusions. For instance, an independent t test requires that the variances in the dependent variable, among the two groups being compared, be relatively equal (not statistically significantly different; homogeneity of variance). It is crucial that researchers take into consideration all of the assumptions inherent in the statistical tests that they employ. It is also important to recognize that increasing the number of statistical tests increases the probability of making a Type I error. The most valid research conclusions are likely to be made by researchers who keep an accurate account of the number of tests they conduct. One common practice is to adjust the significance criterion by dividing .05 by the number of tests conducted.

A practice known as "fishing" also can serve as a threat to the validity of research conclusions. A research study with valid conclusions is likely to have some theoretical basis that serves as a springboard for logical hypotheses to important questions. "Fishing expeditions," in which lots of data are gathered and non-theory-guided statistical tests are employed, are typically easy to spot. They lack tests of logical hypotheses drawn from an existing theory or a theoretical perspective that emerges from previous findings.

Invalid research conclusions also might result from poor reliability of the measures employed, observations made, or treatments implemented. Treatments associated with an experimental condition, for instance, must be administered in a standard fashion. Valid research conclusions are most likely to emerge when each administration of the treatment is as similar to the other administrations as possible and random irrelevancies in the experimental setting are eliminated. Ultimately, reducing treatment administration variance (e.g., conducting each experimental session with the same instructions, the same experimenter, in the same room) will enhance the likelihood of obtaining a true difference between experimental conditions (if one truly exists).

Problems with random selection and random assignment always serve as potential threats to the validity of research conclusions. With the practical costs of highly valid research, random selection is a rare occurrence, contributing to a greater need for replication of research results. The use of within-subject designs can reduce the need for random assignment (because subjects in these designs serve as their own controls). Furthermore, the error variance that results from inadequate random assignment in other designs can be dealt with by controlling for relevant covariates in the statistical analysis.

Other Considerations in Dealing With the Threats to Conclusion Validity

Although statistical issues are highlighted here, researchers engaged in qualitative research are not immune to such issues. Such research might be more likely to violate the "Gricean" maxims of communication and conversation. For instance, it is well known that interview questions can shape a respondent's answers (a type of self-fulfilling prophecy). Subtle changes in the wording used in questions, and even the order of the questions, can have a significant impact on the validity and reliability of the data collected. It is also well known that people do not always behave the way they normally do when they are aware that others are observing them. Essentially, conclusions about significant relationships between variables might be made when in fact they do not exist (and vice versa). It is reasonable to expect that conclusion validity will improve with an awareness of how such confounds might emerge from such research.

It is reasonable to consider the possibility that failure to support a hypothesis, based on a reasonable theory or reliable experimental finding, is the result of invalid research conclusions even when the threat factors (described in the preceding section) are not of concern. Some relationships between variables might be very weak, or they might be very difficult to demonstrate through experimental methods. For instance, asymmetries in perceived similarity are known to exist (e.g., people judge the similarity of a zebra to a horse to be greater than the similarity of a horse to a zebra), but they can be very difficult to replicate in a single experiment. Simply recruiting larger and larger samples might not always result in greater power if the error in measurements also increases with additional subjects.

Besides successfully planning a reliable and high-powered research study, threats to conclusion validity might be addressed in one of three general ways. First, if the design and the nature of the data permit, issues of conclusion validity might be dealt with by arguments that rule out the potential threats. For example, a researcher might rule out some extraneous variables if proper random assignment can be demonstrated. Manipulation checks are also quite useful in this regard. Second, replication, through altering the design of the research manipulations and observations, aids in verifying the conclusions. Observing the same result from a different "angle" can provide greater confidence that the relationship between the variables does or does not exist. Third, employing a different statistical analysis can increase power and thereby reduce the probability of a Type II error (e.g., analyzing data with regression procedures rather than analysis of variance tests).

John V. Petrocelli

See also Construct Validity; Covariate; External Validity; Internal Validity; Power; Significance Level, Concept of; Significance Level, Interpretation and Construction

Further Readings

Chen, H., & Rossi, P. H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning, 10,* 95–103.

Cohen, R. J., & Swerdlik, M. E. (2004). *Psychological testing and assessment* (6th ed.). New York: McGraw-Hill.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings.* Chicago: Rand McNally.

Costner, H. L. (1989). The validity of conclusions in evaluation research: A further development of Chen and Rossi's theory-driven approach. *Evaluation and Program Planning, 12,* 345–353.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York: Academic Press.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54,* 93–105.

Smith, E. R. (2000). Research design. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 17–39). Cambridge, UK: Cambridge University Press.

Trochim, W. M., & Donnelly, J. P. (2006). *The research methods knowledge base* (3rd ed.). Cincinnati, OH: Atomic Dog Publishers.

Variability, Measure of

Variability, meaning differences, is a critical construct in research and science. Objects and events that are constant do not require prediction or explanation. The advancement of science depends on the extent that the differences between objects and events are explainable and predictable. The goal of the scientist and researcher is to create parsimonious scientific models that predict the variability between objects and events and to test those models against the real world. The ability of the scientist and researcher to measure variability allows the assessment of competing scientific models and theories about the world.

There is a confusing array of symbols in the statistical world, all measuring variability. When referring to a theoretical probability model, the variability is symbolized by VAR(X) and might be described with Greek letters. With sample data, variability is measured conventionally using statistics such as the standard deviation, variance, and range. If the measure of variability is the standard deviation or variance, the variability is generally symbolized by the letter s. Measures of sample variance are used as estimates of model variability. Wide varieties of subscripts are used with both model and sample measures of variability to clarify the meaning of the measure. All measures share certain common elements and interpretations.

Theoretical probability models (probability distributions) are mathematical equations used to model distributions of real-world objects or events. In the case of theoretical probability models, variability has a precise definition. Because model parameters are most often symbolized by Greek letters and the variability of a theoretical probability model can often be expressed as functions of these parameters, theoretical model variability is often expressed as equations with Greek letters.

If a sample of tenured full professors was taken and each was asked about the number of hours per week they worked in their academic position, considerable variability would be found in the data. Some professors might work the absolute minimum required by their respective colleges and universities, whereas others might maintain long academic work weeks. The