Encyclopedia of Case Study Research

Reliability

http://dx.doi.org/10.4135/9781412957397.n293

Reliability refers to the consistency and stability of research results and is one of two foundational elements (the other being validity) in conducting rigorous research. Reliability assesses the extent to which the results and conclusions drawn from a case study would be reproduced if the research were conducted again. Reliability in case study research is normally addressed through three techniques: (1) triangulation, (2) interrater reliability, and (3) an audit trail.

Conceptual Overview and Discussion

The concept of reliability is associated with positivist research and addresses the *reproducibility* of results. By contrast, validity assesses the *accuracy* of results. The goal of reliability is to minimize bias and error in the collection and analysis of data to the point that the same results and conclusions would be reached if the research were conducted again.

A common example of reliability is the task of weighing oneself on a bathroom scale. If repeated attempts indicate the same weight, the scale can be said to be reliable. Note that a reliable scale is not necessarily an accurate one: Even though the scale gives a consistent measure, it may indicate a weight that is consistently higher or lower than your actual weight. Thus, reliability can exist without validity, but not vice versa. Put another way, reliability is a necessary but not sufficient condition for validity.

Consistency and stability are two dimensions of reliability. *Consistency* refers to the degree to which the results can be independently re-created within an acceptable margin of error and is a form of measurement error. Consistency can be thought of as the level of variability in the method or instrument of measurement. *Stability* refers to the degree to which the results can be replicated independently at a later point in time and is similar to the replication of an experiment; if the same case were to be re-examined at a later point in time, would the results be the same?

As the use of case studies has gained acceptance within the positivist community, concepts of rigor such as reliability have been increasingly applied to the methodology. However, the importance of reliability in case studies depends to some extent on

the researcher's epistemological perspective. Researchers who adhere to a social constructive or interpretive research philosophy may see case studies as a way to examine a phenomenon embedded within a unique situation at a certain point in time. They may therefore conclude that evaluating reliability is inappropriate, because the research cannot be reproduced.

Application

Reliability in case study research can be assessed by applying three commonly used techniques to address the dimensions of consistency and stability: (1) interrater reliability, (2) triangulation, and (3) an audit trail. These techniques are discussed next in the larger context of consistency and stability.

Consistency

There are two components to consistency: equivalency and internal consistency.

Equivalency

Equivalency is concerned with consistency of observation at a point in time. Case study research is susceptible to error in observation, in particular when a single researcher performs the observation and analyzes the data. In case study research the researcher can be viewed as part of the measurement process. Just as a physical instrument may have error in measurement, so too can an individual in observing or in applying coding or categorization to the qualitative data, introducing bias that **[p. 801 ↓]** impacts reliability. Addressing equivalency requires that steps be taken to minimize the measurement bias of the researcher.

Equivalency can be addressed through the use of multiple researchers who collect and/or analyze the data. Multiple researchers reduce the overall error in measurement by allowing triangulated observations and data analysis that minimizes the error of any one observer. A technique for measuring the equivalence of the researchers'

analyses—interrater reliability—measures the degree to which two or more researchers agree on the application of a judgment scale or coding process. Several approaches to interrater reliability exist, such as kappa statistics (e.g., Cohen's kappa, Fleiss's kappa), correlation coefficients (e.g., Pearson's rho, Spearman's rho), and intraclass correlations. The appropriateness of the individual approach depends on the type of measurement desired.

A common example of an attempt to achieve equivalency is the use of multiple judges during Olympic ice skating competitions. The score assigned to a figure skater's performance is determined by a human judge, whose observations and ratings are potentially influenced by a wide variety of factors—different interpretations of the rules, the judge's country of origin, the style of music being played during the skater's performance, political considerations, and so on. These factors introduce bias and error into the judgment. The use of multiple judges is designed to counterbalance the bias and error introduced by these factors acting on each individual judge.

Internal consistency

Internal consistency refers to the uniformity among similar data points thought to be measuring the same construct. Unlike equivalency (in which the ***measurement method or instrument*** introduces potential bias), in internal consistency data are the potential source for bias and error. For example, suppose a manager is interviewed for his perspective on why the CEO has just been fired. It is possible that the individual simply does not know the true reason, or he or she may have a perspective bias that resulted in inaccuracies in the data collected from the interview.

In case study research internal consistency can be increased by collecting data from multiple sources and by using different types of data, an approach referred to as *triangulation*. Triangulation allows for more confidence in the value of data because the data are derived from multiple perspectives. Triangulation can include the use of multiple *sources*, such as interviewing individuals in multiple departments or at varying levels of management (line workers, supervisors, middle management, etc.) or the use of multiple *data types* (e.g., public documents, such as newspapers, and internal documents, such as memos and e-mails). The measurement principle behind

SAGE researchmethods

triangulation is that the less reliant the data set is on a single *type* of data or a single *source* of data, the more likely that independent researchers would be able to recreate or re-establish the order of occurrence, the degree of influence, or the attitudes and opinions concerning organizational events or characteristics from the past.

Consider the different perspectives of documents generated external to an organization compared with documents generated internally. Documents that are external to the organization—newspaper or magazine articles, government reports, or industry-based promotional material—provide an external representation of facts, figures, and interpretations of events that are gene rally understood and widely available. On the other hand, documents that are internal to the organization—memos, committee or board meeting minutes, company e-mails, or other correspondence—provide internal representations of facts, figures, and interpretations of events that an individual or organization may not necessarily want a general audience to know. Examination of both internal and external documents allows researchers to view data points from multiple perspectives, and this can minimize the bias from any one individual data source.

Stability

Stability represents the consistency of results obtained over repeated measurements and is often measured through test–retest procedures in which a variable is measured at two points in time and then compared to determine whether similar results are generated. In case study, stability depends on whether the case study time line, sequence of events, and changes in the variables under study and their interrelationships across **[p. 802 ↓]** time are repeatable. Because so much of the analysis in qualitative research methods such as case studies relies on researchers gathering, documenting, and inferring variable measurements across multiple data sources, it is vital that the specific process of getting from the raw data to the final evaluations or measurements is made explicit. The absence of an explicit description of the process makes replication by an independent researcher impossible.

An important technique for addressing replication in case study research (and therefore demonstrating the potential for stability) is the *audit trail*—the documentation of the research process, including how and why the data were collected; how the data were

analyzed; and any other decisions or considerations related to the data, the results, or the conclusions that were drawn. Such documentation provides enough detail that another researcher can examine the data collection and analysis process and not only understand what the researcher did and why but also be able to reach conclusions similar to the original researcher's. Even if the nature of the study does not allow a literal replication, the documentation provides a trail that allows for the research—from data collection to conclusion—to be logically replicated.

Critical Summary

Reliability assesses the reproducibility of results and conclusions. Reliability in case study research requires paying attention to both consistency (equivalency and internal consistency) and stability.

There are several techniques researchers can apply to increase the reliability of their research. Using multiple researchers and interrater reliability techniques counterbalances the biases that may be evident when an individual researcher makes observations or analyzes data. Triangulation within and across data sources addresses the potential threat to reliability caused by a lack of internal consistency among data points. Finally, stability can be addressed by documenting the research process so that an independent third party can reproduce the research process from data collection to conclusions.

Kerry Ward and Chris Street

http://dx.doi.org/10.4135/9781412957397.n293
See also

Further Readings

Boyd, B. K., Gove, S., and Hitt, M. A. Construct measurement in strategic management research: Illusion or reality? Strategic Management Journal, (2005). vol. 26, pp. 239–257.

Lincoln, Y. S. Guba, E. G. (1985). Naturalistic inquiry. Newbury Park, CA: Sage.

Merriam, S. B. (1998). Qualitative research and case study applications in education. San Francisco: Jossey-Bass.

Rosenthal, R. Rosnow, R. L. (1991). Essentials of behavioral research: Methods and data analysis (2nd ed.). Boston: McGraw-Hill.

Yin, R. K. (2009). Case study research: Design and methods (4th ed.). Thousand Oaks, CA: Sage.

SAGE researchmethods