Encyclopedia of Research Design

Internal Validity

Internal validity refers to the accuracy of statements made about the causal relationship between two variables, namely, the manipulated (treatment or independent) variable and the measured variable (dependent). Internal validity claims are not based on the labels a researcher attaches to variables or how they are described but, rather, to the procedures and operations used to conduct a research study, including the choice of design and measurement of variables. Consequently, internal validity is relevant to the topic of research methods. In the next three sections, the procedures that support causal inferences are introduced, the threats to internal validity are outlined, and methods to follow to increase the internal validity of a research investigation are described.

Causal Relationships between Variables

When two variables are correlated or found to covary, it is reasonable to ask the question of whether there is a direction in the relationship. Determining whether there is a causal relationship **[p. 620 ↓]** between the variables is often done by knowing the time sequence of the variables; that is, whether one variable occurred first followed by the other variable. In randomized experiments, where participants are randomly assigned to treatment conditions or groups, knowledge of the time sequence is often straightforward because the treatment variable (independent) is manipulated before the measurement of the outcome variable (dependent). Even in quasi-experiments, where participants are *not* randomly assigned to treatment groups, the investigator can usually relate some of the change in pre-post test measures to group membership. However, in observational studies where variables are not being manipulated, the time sequence is difficult, if not impossible, to disentangle.

One might think that knowing the time sequence of variables is often sufficient for ascertaining internal validity. Unfortunately, time sequence is not the only important aspect to consider. Internal validity is also largely about ensuring that the causal relationship between two variables is direct and not mitigated by a third variable. A third, uncontrolled, variable can function to make the relationship between the two other variables appear stronger or weaker than it is in real life. For example, imagine that an investigator decides to investigate the relationship between class size (treatment

variable) and academic achievement (outcome variable). The investigator recruits school classes that are considered large (with more than 20 students) and classes that are considered small (with fewer than 20 students). The investigator then collects information on students' academic achievement at the end of the year to determine whether a student's achievement depends on whether he or she is in a large or small class. Unbeknownst to the investigator, however, students who are selected to small classes are those who have had behavioral problems in the previous year. In contrast, students assigned to large classes are those who have not had behavioral problems in the previous year. In other words, class size is related negatively to behavioral problems. Consequently, students assigned to smaller classes will be more disruptive during classroom instruction and will plausibly learn less than those assigned to larger classes. In the course of data analysis, if the investigator were to discover a significant relationship between class size and academic achievement, one could argue that this relationship is not direct. The investigator's discovery is a *false positive finding.* The relationship between class size and academic achievement is not direct because the students associated with classes of different sizes are not equivalent on a key variable—attentive behavior. Thus, it might not be that larger class sizes have a positive influence on academic achievement but, rather, that larger classes have a selection of students that, without behavioral problems, can attend to classroom instruction.

The third variable can threaten the internal validity of studies by leading to false positive findings or false negative findings (i.e., not finding a relationship between variables A and B because of the presence of a third variable, C, that is diminishing the relationship between variables A and B). There are many situations that can give rise to the presence of uncontrolled third variables in research studies. In the next section, threats to internal validity are outlined. Although each threat is discussed in isolation, it is important to note that many of these threats can simultaneously undermine the internal validity of a research study and the accuracy of inferences about the causality of the variables involved.

SSAGE researchmethods

Threats to Internal Validity

1. History

An event (e.g., a new video game), which is not the treatment variable of interest, becomes accessible to the treatment group but not the comparison group during the pre- and posttest time interval. This event influences the observed effect (i.e., the outcome, dependent variable). Consequently, the observed effect cannot be attributed exclusively to the treatment variable (thus threatening internal validity claims).

2. Maturation

Participants develop or grow in meaningful ways during the course of the treatment (between the pretest and posttest). The developmental change in participants influences the observed effect, and so now the observed effect cannot be solely attributed to the treatment variable.

[p. 621 ↓]

3. Testing

In the course of a research study, participants might be required to respond to a particular instrument or test multiple times. The participants become familiar with the instrument, which enhances their performance and the observed effect. Consequently, the observed effect cannot be solely attributed to the treatment variable.

4. Instrumentation

The instrument used as a pretest to measure participants is not the same as the instrument used for the posttest. The differences in test type could influence the

observed effect; for example, the metric used for the posttest could be more sensitive to changes in participant performance than the metric used for the pretest. The change in metric and not the treatment variable of interest could influence the observed effect.

5. Statistical Regression

When a pretest measure lacks reliability and participants are assigned to treatment groups based on pretest scores, any gains or losses indicated by the posttest might be misleading. For example, participants who obtained low scores on a badly designed pretest are likely to perform better on a second test such as the posttest. Higher scores on the posttest might give the appearance of gains resulting from the manipulated treatment variable but, in fact, the gains are largely caused by the inaccurate measure originally provided by the pretest.

6. Mortality

When participants are likely to drop out more often from one treatment group in relation to another (the control), the observed effect cannot be attributed solely to the treatment variable. When groups are not equivalent, any observed effect could be caused by differences in the composition of the groups and not the treatment variable of interest.

7. Selection

Internal validity is compromised when one treatment group differs systematically from another group on an important variable. In the example described in the previous section, the two groups of class sizes differed systematically in the behavioral disposition of students. As such, any observed effect could not be solely attributed to the treatment variable (class size). Selection is a concern when participants are not randomly assigned to groups. This category of threat can also interact with other categories to produce, for example, a selection-history threat, in which treatment groups have distinct local events occurring to them as they participate in the study, or

a selection-maturation threat, in which treatment groups have distinct maturation rates that are unrelated to the treatment variable of interest.

8. Ambiguity about Direction of Causal Influence

In correlation studies that are cross-sectional, meaning that variables of interest have not been manipulated and information about the variables are gathered at one point in time, establishing the causal direction of effects is unworkable. This is because the temporal precedence among variables is unclear. In experimental studies, in which a variable has been manipulated, or in correlation studies, where information is collected at multiple time points so that the temporal sequence can be established, this is less of a threat to internal validity.

9. Diffusion of Treatment Information

When a treatment group is informed about the manipulation and then happens to share this information with the control group, this sharing of information could nullify the observed effect. The sharing of details about the treatment experience with control participants effectively makes the control group similar to the treatment group.

10. Compensatory Equalization of Treatments

This is similar to the threat described in number 9. In this case, however, what nullifies the effect of the treatment variable is not the communication between participants of different groups but, rather, administrative concerns about the inequality of the treatment groups. For example, if an experimental school receives extra funds to implement an innovative curriculum, the control school might be given similar funds and encouraged to **[p. 622 ↓]** develop a new curriculum. In other words, when the treatment

is considered desirable, there might be administrative pressure to compensate the control group, thereby undermining the observed effect of the treatment.

11. Rivalry between Treatment Conditions

Similar to the threat described in number 9, threat number 10 functions to nullify differences between treatment groups and, thus, an observed effect. In this case, when participation in a treatment versus control group is made public, control participants might work extra hard to outperform the treatment group. Had participants not been made aware of their group membership, an observed effect might have been found.

12. Demoralization of Participants Receiving Less Desirable Treatments

This last threat is similar to the one described in number 11. In this case, however, when treatment participation is made public and the treatment is highly desirable, control participants might feel resentful and disengage with the study's objective. This could lead to large differences in the outcome variable between the treatment and control groups. However, the observed outcome might have little to do with the treatment and more to do with participant demoralization in the control group.

Establishing Internal Validity

Determining whether there is a causal relationship between variables, A and B, requires that the variables covary, the presence of one variable preceding the other (e.g., $A \rightarrow B$), and ruling out the presence of a third variable, C, which might mitigate the influence of A on B. One powerful way to enhance internal validity is to randomly assign sample participants to treatment groups or conditions. By randomly assigning, the investigator can guarantee the probabilistic equivalence of the treatment groups before the treatment variable is administered. That is, any participant biases are equally distributed in the two groups. If the sample participants cannot be randomly

assigned, and the investigator must work with intact groups, which is often the case in field research, steps must be taken to ensure that the groups are equivalent on key variables. For example, if the groups are equivalent, one would expect both groups to score similarly on the pretest measure. Furthermore, one would inquire about the background characteristics of the students—Are there equal distributions of boys and girls in the groups? Do they come from comparable socioeconomic backgrounds? Even if the treatment groups are comparable, efforts should be taken to not publicize the nature of the treatment one group is receiving relative to the control group so as to avoid threats to internal validity involving diffusion of treatment information, compensatory equalization of treatments, rivalry between groups, and demoralization of participants that perceive to be receiving the less desirable treatment. Internal validity checks are ultimately designed to bolster confidence in the claims made about the causal relationship between variables; as such, internal validity is concerned with the integrity of the design of a study for supporting such claims.

Jacqueline P. Leighton

http://dx.doi.org/10.4135/9781412961288.n192
See also

Further Readings

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston: Houghton-Mifflin.

Keppel, G., & Wickens, T. D. (2002). Design and analysis: A researcher's handbook (4th ed.). Upper Saddle River, NJ: Prentice Hall.

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

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton-Mifflin.

SSAGE researchmethods