TOPIC 36 THREATS TO EXTERNAL VALIDITY

Consider this example: You drew a sample from a population and divided it into an experimental group and a control group. To conduct the experiment, you administered the experimental treatment and control condition in a laboratory on a college campus. Suppose the results of your experiment showed that the experimental group improved significantly more than the control group. Can you accurately **generalize** from the sample to the population (i.e., is it accurate to assume that the treatment administered to the experimental group will work as well in the population as it did in the sample)? Also, will the treatment be as effective in the population's natural setting as it was in the artificial laboratory setting? The answers depend on the extent to which the experiment is subject to what researchers call **threats to external validity**.

The first threat is **selection bias** and its interaction with the experimental (independent) variable. You probably recall from Part C of this book that if a sample is biased, our ability to generalize to a population is greatly limited. In a very strict, scientific sense, no generalizations should be made when this is the case. This threat reminds us that if a biased sample of participants is used in an experiment (such as using the students who happen to be in one professor's class), we will not know whether the effects of the treatment (observed in that class) can be expected if the treatment is administered to the entire population. Of course, the way to control this threat is to select the participants for an experiment at random from a population because a random sample is, by definition, unbiased.

Another threat is the **reactive effects of experimental arrangements**. This threat reminds us that if the experimental setting is different from the natural setting in which the population usually operates, the effects we observe in the experimental setting may not generalize to the natural setting. For example, if a treatment is administered in a laboratory to fifth-graders, the responsiveness of the students may be different from the responsiveness of the population of fifth-graders when the treatment is used in public school classroom settings. In other words, it can be risky to generalize from an experimental setting to a natural setting. The way to control this threat is to conduct experiments under natural conditions, when possible.

The possible **reactive effect of testing** (also called **pretest sensitization**) is another threat. This means that the pretest might influence how the participants respond to the experimental treatment. For example, if we give a pretest on racial prejudice and then administer a treatment designed to lessen prejudice, how participants respond to the treatment may be affected by the experience of taking the pretest. That is, having to think about prejudice (while taking the pretest) might change participants' sensitivity to the treatment. This is a problem if we want to generalize about how well the treatment will work in the population *if* the population will not be given a pretest. This threat was discussed in Topic 34, where you learned that we can conduct true experiments without pretests, thereby eliminating this threat.

Multiple-treatment interference is a threat that occurs when a group of participants is given more than one treatment. For example, at first we might give participants no praise for correct answers, then start giving them moderate amounts of praise, and finally give them large amounts of praise. Suppose we find that large amounts of praise work best. Will it work in the same way in the population? Our generalization will be risky if those in the population will *not* first be given the no-praise condition, then the moderate-praise condition, and finally, large amounts of praise. In other words, the fact that these participants first received small and then moderate amounts of praise might make their responses to large amounts of praise different from the responses of a population that will receive *only* large amounts of praise.

Take care to distinguish between **internal validity** (see Topic 35) and **external validity**. As you know from this topic, the external validity of an experiment is concerned with, "To whom and under what circumstances can the results be generalized?" Internal validity is concerned with the question, "Is the treatment, *in this particular case*, responsible for the observed change(s)?" Thus, threats to internal validity are potential explanations for the observed changes. That is, they are possible explanations that become **confounded**[1] with the treatment as an explanation for any observed changes. Threats to internal validity are controlled by using *true experimental designs* (see Topic 34).

[1] *To confound* has many meanings. In this context, it means *to throw into confusion*.

Note that internal and external validity are separate considerations. Even if an experiment has excellent internal validity, it may not be appropriate to generalize the results to other populations because of the threats to external validity discussed in this topic. Likewise, a study with high external validity might have poor internal validity because the threats to internal validity have confounded our understanding of what caused the observed changes. Thus, each one of these two types of threats should be considered and evaluated independently.

EXERCISE ON TOPIC 36

1. Which type of validity deals with the question of whether we can generalize with confidence to a larger population in a natural setting?

2. Which type of validity deals with whether the treatment is responsible for the changes observed in the experimental setting?

3. What is the name of the threat that warns us to be careful in generalizing the results to a population when an experiment is conducted on a nonrandom sample?

4. Suppose a random sample of workers in a factory is exposed to five different reward systems, with each system being used for one month. What is the name of the threat that reminds us that the results may not generalize to the population of workers if the population is to be exposed to only the last reward system tried in the experiment?

5. Suppose an experimental classroom has research observers present at all times. What is the name of the threat that reminds us that the results may not generalize to other classrooms without observers?

6. If a pretest causes a change in participants' sensitivity to a treatment, what threat is operating?

Question for Discussion

7. Briefly describe a hypothetical experiment that has high internal validity but low external validity.

For Students Who Are Planning Research

8. If you will be conducting an experiment, which threats to external validity, if any, will it be subject to? Explain.