TOPIC 35 THREATS TO INTERNAL VALIDITY

Suppose we observe significant changes in participants' behavior in an experiment. Can we attribute these changes to the effects of the treatment(s)? Depending on the design of the experiment, there may be explanations for the changes other than the treatment. These alternative explanations are called **threats to internal validity**. It is easiest to understand them in the context of a poor experiment with no control group (specifically, one in which we pretest, treat, and then posttest one group of participants). Using the symbols described in the previous topic, the design looks like this:

O X O

Suppose the treatment (X) was designed to improve participants' self-concept, and we observe an average gain of 9 points in self-concept from pretest (the first O) to posttest (the second O). Of course, the treatment could be responsible for the increase. Another possibility is the internal threat called **history** (i.e., other environmental influences on the participants between the pretest and posttest). For example, perhaps some of the participants read a new self-help book that improved their self-concepts during the same period of time that the treatment was being administered. Thus, the gain could have resulted from reading the book and not from the treatment.

Another threat in this design is **maturation**. Perhaps the participants matured during the period between the pretest and posttest, and the increase is due to maturation and not the treatment.

Instrumentation is another threat. This refers to possible changes in the instrument (measurement procedure) from the time it was used as a pretest to the time it was used as a posttest. For instance, the particular observers who made the pretest observations may have been less astute at noticing signs of good self-concept than the observers who made the posttest observations.

Another threat is **testing**, which is defined as the effects of the pretest on the performance exhibited on the posttest. For example, while taking the pretest self-concept scale, students may have learned how to interpret the questions. Their posttest performance might be affected by this learning experience.

Statistical regression is another threat that occurs only if participants are selected on the basis of their extreme scores. For example, perhaps a large group of students was administered a self-concept scale, and those in the lowest 20% were selected for treatment in the experiment. A fundamental principle of measurement is that those who are extremely low on a screening test will, on the average, probably have a higher score when tested again, purely because of the nature of random errors created by the less-than-perfect reliability of the measures we use—whether or not the treatment is effective.[1]

The next threat to internal validity can occur when we have two comparison groups that are *not* formed at random. Suppose for example we use the students in one school as the experimental group and those in another as the control group. Because students are not assigned to schools at random, we are using **intact groups** (i.e., previously existing symbols for the groups). We diagram this by putting a dashed line between the symbols for the groups, which indicates that the groups were intact, as shown here:

O X O

O O

Notice that when we do not assign participants to the two groups at random, there is a very strong possibility that the two groups are not initially the same in all important respects, which is the threat called **selection**. Selection can *interact* with all the other threats to internal validity. For example, consider **selection–history interaction**. Because the selection of participants for the two groups was not at random, they may be systematically subjected to different life experiences. For example, it may be that the teachers in the school with the experimental group took a self-concept workshop, which was not part of the treatment, and applied what they learned to their students. Thus, the improvement in self-concepts may be the result of the teachers' efforts and not of the treatment. Another example is **selection–maturation interaction**. Perhaps the two groups, on the average, were at somewhat different developmental stages at the time of the pretest, which would have led to different rates of matura-

[1] Statistical regression is hard to grasp unless you have a good background in measurement theory. However, you may recall from your study of other sciences the principle of "regression toward the mean." Those who are very low will, on the average, tend to be higher on retesting (closer to the mean—an average), and those who are very high will tend to be lower on retesting.

tion in the two groups, which could affect self-concept.

Selection can also interact with a threat called **mortality** (i.e., differential loss of participants from the groups to be compared). For example, those in an experimental school may have a higher dropout rate than those in a control school. If those who drop out have lower self-concepts, the posttest mean for the experimental group will be higher than the pretest mean, resulting in a statistical change in the average that is not the result of the treatment. At the same time, the control group will not have as much of a change because it has a lower dropout rate.

All threats to internal validity can be overcome by using a **true experimental design** (see Topic 34), in which participants are assigned at random to experimental and control conditions. Because random assignment has no bias (or favorites), both groups are equally likely to experience the same environmental events (have the same history), mature at the same rates, drop out at the same rates, and so on.

EXERCISE ON TOPIC 35

1. What is the name of the threat that says taking a pretest may affect performance on a posttest?

2. Suppose that an experimental group is being taught letters of the alphabet as a treatment. At about the same time, the students are watching an educational program on television, from which they learn the names of the letters. What is the name of the threat that this problem illustrates?

3. If observers are more tired and less astute when making posttest observations than when making pretest observations, what threat is operating?

4. What is the name of the threat posed by nonrandom assignment of participants to experimental and control groups?

5. If infants naturally improve in visual acuity and thus perform better at the end of an experiment than at the beginning, what threat is operating?

6. Under what circumstance will statistical regression operate?

7. How can we overcome all the threats to internal validity?

Question for Discussion

8. Suppose a researcher gave a series of wellness workshops over a six-month period and then determined that five of the employees had quit smoking during the six-month period. His or her interpretation was that the workshops caused the decrease in smoking. Is this interpretation flawed? Explain.

For Students Who Are Planning Research

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