

Basic Threats to Internal Validity

Threats Relevant to the Single Group Pretest-Posttest Quasi-Experimental Designs

History—History refers to any external or historical event that occurred during the course of the study that may be responsible for the effects instead of the program itself. For instance, in the 70s when Head Start began, Sesame Street also began airing. Because Sesame Street includes lots of educational information, perhaps it could have been responsible for the apparent effects of Head Start. If we were studying psychotherapy with depressed patients at the time Prozac went on the market, all of our patients might have gotten better by the end of the study, because they went to a psychiatrist and got a prescription for Prozac.

Maturation—A second potential explanation for results in these designs is maturation. Maturation refers to a natural process that leads participants to change on the dependent measure. For instance, perhaps Head Start kids start getting better just because they were getting older during the study. One classic problem with many health studies is that patients get better by themselves (e.g., headache medicine trial). Maturation can refer to a decline as well as an improvement. For instance, some loss of short term memory abilities as you get older, or decline in girls' math scores and boys verbal scores at around age 12 or 13. You can distinguish between history and maturation because maturation is internal, a natural course of things having to do with some quality of the participants in the study. History has to do with an external event of some kind.

Instrumentation—Instrumentation refers to an improvement or decline that results from the measure. This is most commonly found with observational measures. There is a common problem with observers getting better at observing. As an example, say we were observing young kids' verbal behaviors in Head Start. It may be that the observers miss a number of behaviors indicative of good verbal skills during the pretest, but they were more likely to count these behaviors at post-test. Instrumentation can also be a decline.

Testing—A threat that involves an improvement in scores on the post-test due to taking the test a pretest is called "testing." People commonly improve on standardized tests such as intelligence tests, SATs, or GREs. Something about taking the test the first time leads to a change in the test the second time, such as learning the answers or how the trick questions are set up. Perhaps the Head Start students learned the types of verbal questions (e.g., analogy) on the pretest and therefore knew how to do them better by posttest. In this case, it would not be the treatment that had an effect but the experience of taking the test once that led to an improvement.

Mortality (Attrition)—Mortality refers to people dropping out of the study during its course. For instance, Head Start children having the most difficulty drop out of the program and by the end of the study the participants who remain have higher academic skills on average. I wonder whether the recent trends of more frequently expelling students from public schools will help improve a school's test scores.

Regression to the Mean—Regression to the mean is the most difficult to understand. It has to do with a sort of statistical fluke. Whenever scores fluctuate over time for any reason, extreme scores tend to move toward the middle, and middle scores tend to move toward the extreme. Regression to the mean is most likely to be a problem in a study in which participants are chosen for their extreme values. For example, an early study evaluating the effects of Sesame Street indicated that it was especially helpful to the most disadvantaged kids. The kids with the lowest skills improved the most. This could have been because the educational material that the Sesame Street presents was the most helpful to the kids who knew the least; or because those who did especially poorly on the pretest had lower scores by chance, improvement was simply a matter of random changes back toward their true scores.

Threats Relevant to Two-Group Quasi-experimental Designs

Selection bias. Selection bias refers to any difference between the groups before the start of the study. Two common ways selection bias can occur is through self-selection or experimenter-selection. **Self-selection** bias is when the participants themselves choose which group they are in. In the Head Start example, parents that choose to be in the program, may have more motivation to teach their kids certain skills. In this example, kids improved because, motivated kids or families choosing to be in Head Start were more motivated at the outset than those who chose not to be in the program. **Experimenter selection** bias is when the researcher chooses who is assigned to each group (e.g., program vs. control group). In the Head Start study, for instance, because of scarce resources, there may only be a limited number of slots available in the program, so the families that contact the program first are assigned to the program group and the families who contact the program last are assigned to the control. The families that contact the program earliest may be more motivated to educate their kids, and, therefore, the kids do better by the end of the study—not because of the effectiveness of the program but because of initial differences between the characteristics of the program and control groups. Because the researcher assigned families to the groups in a way that selected more motivated families, the results are due to experimenter selection bias.

As long as we start out with comparable groups initially, the previous threats to internal validity are not a problem. It's very difficult to ensure that groups are completely comparable initially, however. The general weakness of two group designs is selection factors, but selection number of different forms. All of the following represent a *differential change* in the groups as a result of selection plus other threats.

Selection & History—Head Start kids are more motivated and more likely to watch Sesame Street when it came on.

Selection & Maturation—Head Start kids are more motivated and thus more likely to develop skills faster than those in control.

Selection & Testing—Head Start kids are more motivated so they are more likely to figure out the tricks to the math questions. Consequently, with a second testing, they improved.

Selection & Instrumentation—Perhaps observers in control group get bored and less likely to pick up on kids improvement, so it looks like the program kids do better.

Selection & Mortality—In the control group, the most motivated kids and families are more likely to drop out of the study because they found another preschool opportunity. This leads to differential attrition in which the kids most likely to improve will leave the control group. At posttest, the treatment group will be better than the control group.

Selection & Regression—Kids who are the furthest behind or have the most disadvantaged families are assigned to the treatment group. Because they are more extreme to begin with, they show greater improvement due to regression rather than program effects.

Further readings on design

- Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design & analysis issues for field studies*. Boston, MA: Houghton Mifflin.
- Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Campbell, D.T., & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: RandMcNally.
- Reynolds, K.D., & West, S.G. (1988). A multiplist strategy for strengthening nonequivalent control group designs. *Evaluation Review*, 11, 691-714.