Learning Guide: True & Quasi-Experiments

If you have forgotten key ideas about sampling, review key concepts before you try to prepare for this module. View the slide shows about True Experiment Requirements and Quasi-Experiments and my document Types of Experiments. We have little information about the effectiveness of many interventions, outside medicine and to a lesser degree education, because we fail to use true experiments to evaluate most intervention programs. We simply do not know if the intervention actually produces an effect. True experiments rely heavily on random assignment to ensure that the variance (different characteristics) of study participants do **not** affect the outcome of an experiment. We know people differ in many ways. We know we cannot screen for all of these differences. We know that some differences may affect the outcome of the experiment – even when we screen for differences that we know or suspect could affect the outcome. We use random assignment to try to "randomly distribute" these unknown differences between treatment and control groups. As a result, many consider the true experiment to be the only "gold standard" for showing that the interventions that we develop really work. Others argue that other kinds of designs are as good or better for evaluating the effectiveness of interventions. I would argue that experiments (true and quasi) are critical to showing direct cause and effect and that we should incorporate the other design groups to extend and expand our understanding – not abandon them. This is, of course, a result of my scientific realist epistemology.

1. To whom and why is it important that evaluation evidence is credible?

2. Some argue that the kind of evidence provided by RCTs is the most credible evidence that an intervention "really works." What characteristics of the true experiment support this perspective? (Hint: look at threats to validity)

3. Many factors are important in evaluating a body of knowledge (not just one study, but the entire body of knowledge). Proponents of different approaches to evaluation of impacts stress different kinds of impacts. To some degree at least, proponents of the RCT in evaluation stress the importance of **average program effects** whereas proponents of other approaches stress other kinds of evaluation questions. What are some of the other kinds of evaluation questions that can be posed?

4. As we discussed in sampling, the need to **define what constitutes a meaningful effect** lies at the heart of the discussion about the value of RCTs and other designs. As a scientific realist, I would argue that the contributions of RCTs are critical, but that you, as a researcher, must consider context for the evaluation and clearly assess what constitutes a meaningful effect size, as opposed to a significant effect size. Think about any intervention of interest to you. Do you anticipate large or small effects? How would this expectation influence your choice of what kind of evaluation evidence needed?

5. Another argument against RCTs is that these designs are meaningless because context and the traits of people (individual variance) inevitably affect the results of a study. Why would a scientific realist reject this argument as a **generally valid argument** against RCTs as a form of evidence?

6. Refer to the Green & Glasgow article. Those authors say "*Starting with the proposition that "if we want more evidence-based practice, we need more practice-based evidence."* Their article offers three kinds of advice about ways to improve evidence-based practice and

practice-based evidence. (1) They offer questions and guides that practitioners, program planners, and policy makers can use to determine the applicability of evidence to situations and populations other than those in which the evidence was produced (generalizability). (2) They provide some procedures that practitioners can use to decide to what degree and how to adapt evidence based interventions (the results of experiments) to the conditions in which they work *without compromising the critical aspects of the experimental protocol.* (3) They suggestion criteria that other researchers and practitioners can use to evaluate external validity and potential for generalization. Give two or three examples of each of these types of recommendations that you think are especially relevant to the research and practice that is of interest to you.

7. There can be a big difference between the planned intervention and the actual intervention that is implemented. "Stuff happens," especially when interventions are implemented under relatively uncontrolled conditions (like our experiment in Haiti). How does lack of fidelity between conceptualized and actual implementation of a treatment affect internal validity? External validity? Explanatory power?

8. Explain how sample size, and specifically power, affects the researcher's ability to detect differences between treatment and comparison groups when the interventions are relatively similar in basic approach.

9. There are at least two ways to improve ability to detect small, but significant, differences between outcomes, increasing power (bigger sample) and increasing the homogeneity of the sample. However, many experiments rely on relatively small samples, often because it is simply impossible to either (a) find or (b) manage a larger number of participants. Explain what this means in terms of defining the theoretical population and the implications for external validity.

10. What is independent replication? Why is independent replication of studies an important way to increase the internal and external validity of the body of knowledge that results from experiments (true or quasi)?

11. Many interventions are complex – they have many components. Why does this make it difficult to understand which components are most important or critical? What can researchers do to help uncover which components are the most critical? Answer from a design perspective – components in the design that you could use to better "sort out" the relative effect of different aspects of the treatment.

12. There are three absolute requirements for establishing any kind of causal effect. (1) Cause must precede effect in time. (2) Cause must co-vary with effect – the stronger the cause, the stronger the response (or vice versa for negative covariation). For example if we have three levels of a treatment (low, medium and high), we should expect the biggest effect from the "high" treatment, and all treatments (causes) should produce a greater effect than the control. (3) The researcher must eliminate logically and empirically alternative causes for the observed effect. In reality, eliminating every possible alternative cause that one could conceivably imagine is very difficult, if not impossible (although positivists claim this is the goal). As a scientific realist, I would argue that alternative explanations for the effect should be highly unlikely." Only one of these three requirements is necessarily more problematic for quasi-experiments than for experiments, the third one. Why does the failure to assign participants to treatment and control groups randomly make it hard to meet this requirement?

13. Compare the degree to which a true experiment and a quasi-experiment help researchers avoid each of the potential threats to *internal validity* and to *external validity*. Look at those lists and think about this as "eliminates or nearly eliminates threat," "reduces the threat," and "has no effect on reducing the threat."

14. Whatever design, true experiment or quasi-experiment (or others, as we will see), what does "matching" samples mean? Why is this hard to do, or at least do well enough to reduce major threats to internal validity?