Strengthening Quasi-Experimental Designs

Quasi-experiments are subject to threats to both internal and external validity because the random assignment requirement of the true experiment is missing. However, there are many ways that you can strengthen internal and external validity through design and the use of statistical approaches to eliminating non-treatment effects on the outcome. Gersten et al. and Shadish & Cook give many suggestions. I focus here on a three more examples that involve using the full array of quasi-experimental designs possible and, in the third example, statistical control of variance. I give three examples here. Explore the designs in "types of true experiments" to see how you can adapt those designs to strengthen quasi-experiments. Also think about using statistical controls – like replication over time, replication over units, and ANCOVA.

Non-Equivalent Group Designs

One of the most common reasons for going to a quasi-experiment is that your potential test subjects are naturally embedded in some larger groups such as students in classes, hospital patients in wards, monkeys in troops, youth in 4-H clubs. You cannot randomly assign participants to comparison groups.

Example

There is a new set of experiential learning materials about environmental education for high school students. You believe that this way of learning about the environment is more apt than traditional teaching methods to produce change in environmentally responsible behavior, behaviors like recycling, turning off the lights, not letting the water run while you brush your teeth. You get three of seven science teachers in a school district to agree to try the new approach. The other 4 teachers will use the old approach. Your participants are students in their classes. You will measure environmentally responsible behavior (how often they recycle, etc.) before and after the educational intervention, new or traditional. Here are the results, which show a very significant difference between treatment and comparison groups.

Group	Pre-test score	Post-test score	p-value
Treatment (new approach)	18	43	0.02
Comparison (traditional approach)	22	31	

This looks like a great result, what you predicted. The problem is that your test subjects, the students, come in **groups**. You did **not** assign Student A to comparison group and Student B to treatment group. The big problem with ***all non-equivalent group designs*** is that you cannot be sure that the comparison and treatment groups were equivalent in every regard that could affect the outcome. For example, what if the three teachers who agree to try the new approach are just better teachers than the other four (more enthusiastic, more willing to help students, more innovative, etc.). In other words, teacher quality could have had an effect on the outcome. Perhaps students in the classes of the three teachers who were willing to try your new approach would change their behavior more, no matter which approach the teacher used. The very fact that only three teachers of seven were willing to try your innovative approach may be an indicator that they are the "good teachers."

Which of these threats to internal validity increase, compared to a true experiment, in this non-equivalent groups design? Explain.

- History
- Maturation
- Testing
- Instrument Decay
- Regression
- Mortality
- Selection Bias
- Selection Interaction

Which of these threats to external validity increase?

- Selection interaction
- Sensitization
- Artificial response
- Theoretical adequacy

What could you do to reduce the threat from the "good teacher" factor in your quasi-experiment? Consult the Statistics Guide – look at the ANCOVA design. You could use some surrogate measure of how students in each of the seven teachers' classes perform ***in general*** to account for this influence – something like FCAT scores in science. If there is a "good teacher" effect, students in the classes of the three teachers who tried your approach should have ***generally higher*** scores on science tests than students in the other teachers' classes. You could use these scores as an independent variable to "extract" the teacher effect.

Switching Replications with Non-Equivalent Groups

Sometimes a different design other than the "classic" form of the experiment can increase the strength of quasi-experiments.

Example

You want to give people a new anti-malaria drug in Ngandajika, Congo. You can't very well go to a village and say "Well, tough luck, you're in the control group. No drug for you." People know each other. They will share drugs, adults will give theirs to kids, and other things well beyond your control will happen. You pick two villages. You try to pick equivalent villages in terms of non-experimental characteristics that could affect the experiment (village size, average age of residents, distance from a river – characteristics that would cause malaria to be more prevalent

in one village than the other). You measure malaria parasite levels in the blood in both villages. You give people in one village the drug and not in the other. Then you measure outcome (malaria parasites in the blood).

Group	Pre-Test (titre)	Post-Test (titre)	Difference pre- and post-test	p-value for difference between groups
Treatment (drug)	61	56	-5	0.03
Control (no drug)	53	49	-4	0.04

Your drug did not work! However, two things bother about the results. First, you see that the parasite load went down in **both groups.** This causes you to wonder if something else (dry weather?) caused the drop. Second, perhaps of more concern since you might be able to statistically control for rainfall, the control (no drug) group lower parasite levels before the experiment was conducted. ? This is called a pre-test advantage. Maybe your drug did not work because the initial malaria parasite level was higher in the treatment than control village? Would it have worked had the treatment village had the same pre-test malaria parasite blood load as the control? You can't be sure and finding two villages with identical average parasite loads is highly unlikely. The switching replications design would eliminate most of the concern.

Switching Replications with Noequivalent Groups (just like switching replications in the true experiment)

Village	Sequence of Treatment & Measurement				
1	Pre-test	Treat with drug	Post-test 1	Withhold drug	Post-test 2
2	Pre-test	Withhold drug	Post-test 1	Treat with drug	Post-test 2

In this design you have some control over the differences in malaria parasite load in different villages because all villages get both the treatment and the control. Which of the threats to internal validity does this design reduce? Why? Which of the threats to external validity does it reduce? Why?

Separate Pre & Posttest Groups

Sometimes you cannot keep the same subjects in an experiment for long enough to make comparisons between treatment and control. Separate pre and posttest group designs measure the same variables for **different groups** of participants over time.

Example

The local school board will implement a new SAT preparation program starting in next year. You want to know whether it improves SAT scores. The problem is that, like many government interventions, ***every student*** will get the treatment. There's no way to set up a true control group. The separate pre and post-test design can make it possible for you to evaluate this new program. You have the SAT scores of students over the past several years when there was no SAT preparation program. This is your control group. After the new program is started, you can get the SAT scores of the new students who will get the program. This is your treatment group. You can compare them with something like a t-test, but you are making many assumptions that

nothing else changed in the schools – just the intervention, the preparation program. There are several steps you can take to make your design stronger. Compare the average scores for each school using a two-way ANOVA. This will help eliminate differences due to “school environment” or “school quality.” Second, replicate for several years. For example, you might compare the SAT scores for three years before the program and for the first three years of the program. This will help eliminate “yearly variance”. A factorial ANOVA will allow you to do this. You would have two treatments (program and control), school (say 10 of them), and replications (3 years). If you use different pre and post-test groups as a design, statistical control for all those things that can affect the outcome becomes very important.