Pre-lecture Notes IV.2 – Pre-test / Post-test Designs

Imagine that you are a clinician (of some sort) and you have developed a new therapy to "fix" a problem. (For example, imagine that you are a clinical psychologist and you have developed a new way to help clients cope with excessive anxiety.) As an empirical scientist, it is not enough to have a well-thought-out theory behind your new therapy; you need actual evidence that it really does work. The question is how will you get this evidence.

The most popular experimental design for this type of research is the two-group, pre-test/post-test design. One group of subjects will receive the new therapy; the other will not. (Note: when all you are trying to do is show that your new therapy works, the second group receives no treatment at all; when you are trying to show that your new therapy is as good or better than an existing therapy, the second group receives the existing treatment. For now, I'll assume that the second group receives no therapy.) This aspect of the design is clearly a between-subjects manipulation, because the two conditions – namely, your new therapy vs nothing – are experienced by different groups. On the assumption that your treatment is designed to have a long-lasting effect, you really can't do this comparison within-subjects, because once a subject has received the new therapy, they are changed forever. (Plus, the demand characteristics of a within-subject comparison between therapy and no-therapy would be so huge that no-one would even try to do this within-subjects.)

Now, therapies are supposed to cause a change in the behavior (or feelings, etc.) of people, which is why you take one measure of the issue *before* the therapy and another measure *after*. The *before* measure is the pre-test and the *after* measure is the post-test. These two measures are clearly within-subjects. So, in total, this design has one between-subjects factor – namely, therapy vs no-therapy – and one within-subjects factor – namely, *before* vs *after*.

The general idea is that an effective therapy will induce a beneficial change between the pre- and post-tests. The control group – i.e., the no-therapy group – is there to get a measure of how much the problem issue would change (e.g., "fix" itself) just due to the passage of time. The hoped-for data is that the (beneficial) change between the pre- and post-tests for the group of subjects receiving your new therapy is larger than that for the other group. That would suggest that your treatment is better than nothing.

One option, therefore, is to calculate a "change score" for every subject during the pre-processing of the data. In other words, every subject (from all conditions) ends up contributing a single piece of data: how much they "got better" during the course of the experiment. You could almost think of this as being a paradigm measure: the operational definition of "therapeutic effectiveness" is the amount of change between the pre-test and the post-test. This is actually the most-popular way to analyze the data. It's very easy and straight-forward because it boils it all down to an independent-samples ***t***-test between the two groups (i.e., therapy vs nothing) using the change scores as the data.

Unfortunately, this is not the right way to analyze the data. To see why, think about the key threat to the internal validity of a between-subjects design. (If you can, always try to avoid being distracted by the details of the theory and/or situation, and always return to the basic issues from the first three sections of the course.) The key threat to internal validity that is specific to between-subject designs is a failure of random assignment: i.e., that the two groups of subjects were not equal to begin with, so any finding of a

difference between the groups in the data (at the end) might be due to pre-existing differences and not due to the conditions created by the manipulation.

What were our options for dealing with the possibility of a failure of random assignment? They were to use matching, add a verification measure, or to add a covariate, with preference going to the third. This is exactly what we can do here: use the pre-test as the covariate and, thereby, equalize the groups statistically. This is what should be done and in lecture we'll discuss how and why.