



Compared to What? Using Nonequivalent Pre-Post Control-Group Designs

It is important to measure outcomes for the population affected by your project, but measuring outcomes in isolation does not tell the full story. You need to consider how the outcomes may have been different if your project had not been implemented. That is, you need to know, “compared to what?” This brief focuses on using nonequivalent pre-post control-group designs and represents the third brief in a five-part series, “Compared to What? Identifying Good Comparison Data to Assess Project Results.”

Using a pre-post design will provide stronger evidence of your project's effectiveness than a post-test alone.

In general, pre-post designs refer to evaluation or research designs in which participants are administered some type of assessment before and after the project is implemented. Measures may include knowledge assessments, observations, and surveys. Typically, the same measures (or a different version of the same assessment) are used for the pre-test and the post-test, and changes in scores from the pre-test to the post-test are interpreted to reflect the effectiveness of the project. The pre-test is important because it provides necessary information about where participants started, enabling you to examine how participant performance changed as a result of your project.

There are several types of pre-post designs, all of which provide stronger evidence of your project's contributions than a post-only design. And, a nonequivalent pre-post control group design will provide stronger evidence than a pre-post design without a control group. This design can be used to demonstrate your project's effectiveness by showing changes in outcomes over time and allowing you to compare results for individuals who did and did not participate in your project. In the nonequivalent pre-post control-group design, you select a control group at the outset of your project. Before your participants begin to work with the project, you administer a pre-test to both the participant and control groups. Then your participants engage as planned with your project while the control group does not. At the end of your project, both groups complete a post-test. Participants are not randomly assigned to groups, which is why the groups are “nonequivalent.” The requirements involved with evaluations that use random assignment are often too resource intensive or expensive given project funding levels.

The addition of the control group makes this a more robust design than the one-group pre-post design. In fact, the nonequivalent pre-post control-group design is referred to as a “quasi-experimental” design. Because you are analyzing the changes from pre-test to post-test in both the project and control groups, changes due to maturation, history, or testing effects are more likely to be the same across both groups; therefore, you can more confidently assume that pre-test to post-test differences in gains between your project and control groups are a result of your project activities.

INTERNAL VALIDITY

Internal validity assesses whether the results of an intervention are due only to the variable being studied or if other factors might have influenced the outcomes. In the context of pre-post designs, internal validity refers to the extent to which any changes seen from the pre- to post-tests are likely attributable to your project. Specific threats to internal validity include maturation (changes that occur naturally over time as a result of participants' experience), testing effects, participant history, and events that may or may not be known that occur between the pre- and post-tests.

If you have the opportunity to use random assignment, you can eliminate, or at least minimize, the risk of selection bias. (Please refer to CIPP's [Evaluating Special Education Programs Resource Toolkit](#) where you can find additional information about random assignment as well as other evaluation designs.) Another way to reduce selection bias, if random selection is not feasible, is matching. Matching involves creating groups of individuals or other units with similar scores on one or more variables (e.g., school size, ethnicity) that are presumed to be related to the outcomes of the planned intervention. If properly done, the treatment and control groups have the same (or very similar) characteristics on those variables. An important step in the use of matching is calculating baseline equivalence on your pre-test measure. Essentially, baseline equivalence involves establishing the differences in pre-assessment means between your treatment and control groups. You will use these calculated differences to adjust your comparison analysis when you assess post-intervention outcomes at the end of your evaluation. We recommend working with a statistician to perform this procedure. If you need additional information, the [What Works Clearinghouse](#) standards include methods for establishing baseline equivalence in nonequivalent pre-post control-group designs.

AN EXAMPLE OF A NONEQUIVALENT PRE-POST CONTROL GROUP DESIGN

As part of a Personnel Development Program (PDP) project evaluation, faculty rated student performance during their practicum using the Teacher Rating and Assessment Instrument for Teachers of Students with Significant Disabilities (TRAIT-SD). The TRAIT-SD uses classroom observation, teacher interview, and document review to rate a teacher on 37 specific skills. The faculty used the TRAIT-SD to rate the scholars participating in the PDP project at the beginning and end of their practicum experience to document gains over the course of the school year. At the beginning and end of the school year, the faculty also used the TRAIT-SD to observe a cohort of scholars who did not participate in the PDP project. Scholars were matched on key demographic variables and their current GPA. Baseline equivalence on the pre-assessment was also calculated. When analyzing the data, the faculty adjusted the changes in scores from pre-assessment to post-assessment for the treatment and control groups and compared them to determine if the PDP project contributed to significant student gains on the TRAIT-SD.

ADDITIONAL RESOURCES

This brief is part of a series, “Compared to What? Identifying Good Comparison Data to Assess Project Results.” For additional information on evaluating special education programs more generally, you may wish to consult the [Evaluating Special Education Programs: Resource Toolkit](#) available on the [OSEP IDEAs That Work](#) website. To learn more about identifying good comparison data, you may wish to refer to the other briefs in this series, “An Overview: Identifying Good Comparison Data to Assess Project Results,” “Using One-Group Pre-Post Designs,” “Using Single-Case Interrupted Time Series Designs,” and “Using Extant Data.”