Compared to What?
Using Single-Case Interrupted Time Series 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 single-case interrupted time series designs and represents the fourth 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 measure) are used for the pre-test and 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. A single-case interrupted time series design is particularly strong because it will allow you to plot trend lines before and after project activities, thus documenting changes over time. Single-case designs (also known as single-subject designs) involve in-depth study of a single person, group, or institution. A time series design measures the same outcome for your project participants multiple times before and after project implementation, and "interrupted" refers to the start of your project activities. Ideally, the pre- and post-assessments are numerous (eight of each is a common standard) and the intervals between assessments are fairly long (up to 1 year). For project evaluation purposes, however, these metrics are seldom realistic. If your intervention has a strong effect, you might be able to conduct fewer assessments (perhaps four pre-assessments and four post-assessments) at shorter intervals (perhaps once a month). This is sometimes referred to as an abbreviated time series design. We recommend you consult a methodologist with expertise in this area to determine how many assessments to use for your project.

You can plot data on a graph and examine the trends before and after project activities. The patterns can be studied for level, trend, variability, consistency, and percentage of overlap. In addition, you can use regression modeling. History and selection bias remain threats to internal validity with this design. With regard to history, another large event that overlaps with your project activities (e.g., a new reading curriculum) may threaten the validity of your evaluation findings because you can't distinguish between the effects of your intervention and the co-occurring event. If the intervals between assessments are relatively long, selection bias becomes an issue as well; the population of interest may shift in meaningful ways as people come and go over the course of data collection, diminishing the strength of the conclusions you can draw.

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.

One potential way to control for threats to internal validity is to replicate your project activities in several schools, districts, or states (in different locations and, ideally, at different times), using an interrupted time series design each time. If you see the same trend across locations and times, you can be more certain that any changes in the trend lines are due to your project activities. Another option that you can use to control for history and selection bias, if the population size allows, is to add a control group and collect data at the same time points as the project group. The trend data for the control group can help you understand what the trend lines would have looked like without your project activities, and, compared to a design using a single project group, you can more confidently attribute any differences to your project activities.

AN EXAMPLE OF AN ABBREVIATED INTERRUPTED TIME SERIES DESIGN

In three districts in each of three states, a Technical Assistance and Dissemination Center provided intensive technical assistance focused on reducing the rate of suspensions and expulsions in schools with high rates for students with disabilities. The evaluation team decided to use an abbreviated interrupted time series design to examine the effects of training offered as part of the technical assistance. The training sessions were implemented in a staggered fashion in each state, with one district in each state being trained in year 1 of the project, a second district in year 2, and the third in year 3. The training always occurred midway through the school year. Because monthly discipline data was available for each district, the evaluation team could obtain the number of suspensions and expulsions for students for 4 months before and 4 months after the training. The fact that implementation was staggered and in different locations, helped control for history and selection bias. The trend lines were compared across districts. Regression analyses were conducted to determine the significance of changes in trend lines from before to after the training sessions.

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 and Using Good Comparison Data to Assess Project Results," Using One-Group Pre-Post Designs," "Using Nonequivalent Pre-Post Control-Group Designs," and "Using Extant Data."