Compared to What?
Identifying Good Comparison Data to Assess Project Results

CIPP

It is important to measure the outcomes of 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 the project had not been implemented. That is, you need to know, "compared to what?" Comparison data can dramatically strengthen your evaluation by helping you tease out the unique contribution of your project to its expected outcomes and by describing how your project has influenced outcomes for the intended recipients. Collecting comparison data can help to bolster your evaluation findings, even if the use of appropriate comparisons is limited to one of several intended outcomes or one aspect of a complex project or intervention.[1]

This overview introduces a series of briefs on using comparison data to evaluate projects funded by the U.S. Department of Education's Office of Special Education Programs (OSEP). Specifically, three briefs focus on pre-post designs that are commonly used in OSEP evaluations: one-group pre-post designs, nonequivalent pre-post control-group designs, and interrupted time series designs. A fourth brief addresses using extant data sources to enhance comparisons. These briefs are meant to introduce the topics to stimulate interest and generate ideas, not to serve as "how-to" guides. This introduction briefly describes the two broad topics covered in the briefs and then notes some important additional considerations.

Compared to What? **Using One-Group Pre-Post Designs**	Compared to What? **Using Interrupted Time Series Designs**
Compared to What? **Using Nonequivalent Pre-Post Control-Group Designs**	Compared to What? **Using Extant Data**

TOPIC 1: COMPARED TO WHAT? PRE-POST DESIGNS

Pre-post designs are evaluation designs in which your project's participants complete some type of assessment before and after implementation of your project (or specific project activities). Assessments can include knowledge assessments, observations, or surveys, and the measure used to conduct the pre-test and post-test is usually the same (or very similar). Participants' knowledge, skills or behaviors before participation are compared to their knowledge, skills or behaviors after participation in your project—and you would interpret any changes from pre- to post as a potential reflection of your project's contributions.

---

[1] In this case, an intervention refers to a product, service, or education program.

Below, and in separate briefs, we highlight three common types of pre-post designs: one-group pre-post designs, nonequivalent pre-post control-group designs, and interrupted time series designs. However, there are various other pre-post designs that you might use, and in selecting the appropriate design for your evaluation, it is important to consider the level of rigor of the design and the resource requirements for implementing it. For example, a randomized controlled trial is often referred to as the gold standard of rigor, but the resources required to implement a successful randomized controlled trial are substantial. Below, we visually display the level of rigor and resources to implement several quasi-experimental designs, all of which are detailed in the **Evaluating Special Education Programs: Resource Toolkit.** The designs described in the briefs that follow are circled in the figures below.

Topic 1a: One-Group Pre-Post Designs

The **one-group pre-post design** is the simplest pre-post design: Participants are measured before and after implementation of your project activities (or a subset of activities), so there is a point of comparison and changes can be assessed. While the one-group pre-post design is often the least resource intensive of the pre-post designs to implement, it is also the least rigorous. More rigorous design options include the **nonequivalent pre-post control-group design** and the **interrupted time series design.**

PRE-POST DESIGNS: RIGOR AND RESOURCES
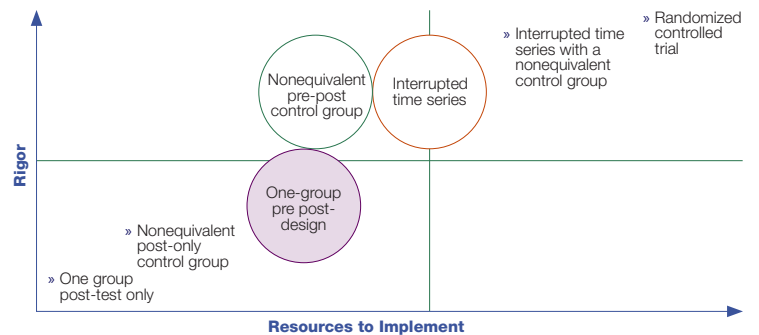
DROPOUT PREVENTION PROGRAM EXAMPLE: ONE-GROUP PRE-POST

An OSEP Technical Assistance and Dissemination Center plans to provide intensive technical assistance on a dropout prevention program for students with disabilities. One of its evaluation questions is, "Did intensive technical assistance on dropout prevention improve student attendance and academic achievement in participating districts?" How might this center use a pre-post design to answer this question?

One way would be to collect data on the outcomes of interest before and after the center provided technical assistance. Changes in student attendance and progress monitoring scores could be analyzed pre- and post-program, which would allow the evaluation team to have some indication of changes the project may have generated. Evaluators could follow up this pre-post data collection with qualitative interviews to determine whether project participants attribute the changes to the project.

Topic 1b: Nonequivalent Pre-Post Control-Group Designs

The **nonequivalent pre-post control-group design** is similar to the one-group pre-post design, except that it includes a comparison group. In this design, the treatment and control groups are not randomly assigned to conditions. Because your comparison group must participate in the pre-assessment along with project participants, it is important that you select your comparison group before implementing project activities. Post-participation data can then be collected from both groups.

For example, an OSEP-funded Parent Training and Information Center (PTI) offered an Individualized Education Program (IEP) clinic for families. One of its evaluation questions is, "Did attendance at an IEP clinic improve parents' knowledge of their rights?" How might this center use a pre-post design to answer this question? Evaluators might survey parents before and after the clinic, asking them about their knowledge of the Individuals with Disabilities Education Act (IDEA) and how well they felt they could communicate their concerns and ideas during IEP meetings. To document the outcome of the clinic, the evaluators might also conduct a similar survey, in the same timeframe, for a sample of parents who contacted the PTI but did not attend the IEP clinic and compare the results over time for the two groups. The findings would indicate if the parents who attended had a better understanding of IDEA than parents who did not attend, giving an indication of whether the clinic helped to improve parent understanding.
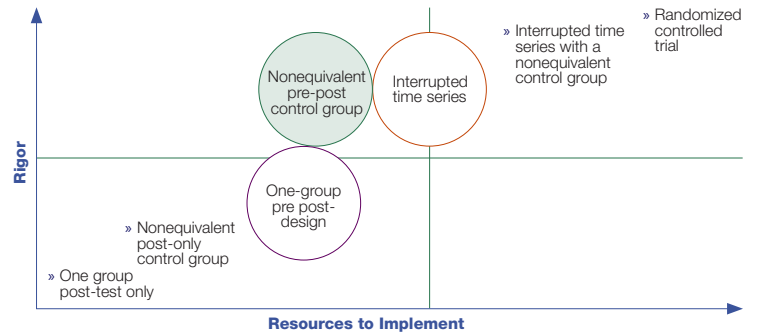
DROPOUT PREVENTION PROGRAM EXAMPLE: NONEQUIVALENT PRE-POST CONTROL-GROUP

The OSEP-funded Technical Assistance and Dissemination Center on dropout prevention described in the previous example could make its pre-post design more rigorous by using a nonequivalent pre-post control group design to answer the same evaluation question: "Did intensive technical assistance on dropout prevention improve student attendance and academic achievement in participating districts?"
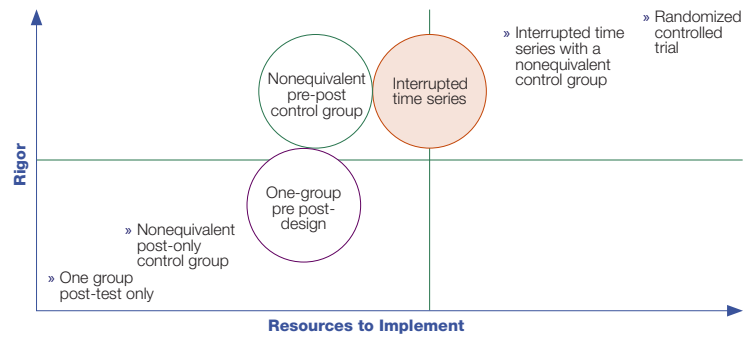
The evaluators could add a control group of districts or schools that did not receive intensive technical assistance to their original pre-post design. They could collect the same data from both groups: student attendance and progress monitoring scores, before and after the program. Changes in the measures could be analyzed pre- and post-intervention, and then the changes in scores from pre to post could be compared for both groups. If the students in districts who received technical assistance had significantly greater increases in their attendance or achievement than those in districts that did not, it would strengthen the argument that the changes are attributable to the dropout prevention program. Evaluators could also follow up this pre-post data collection with qualitative interviews to determine whether project participants attribute the changes to the project.

Topic 1c: Interrupted Time Series Designs

An **interrupted time series design** is similar to the one-group pre-post design in that it doesn't have a control/comparison group. However, the design is more rigorous than the one-group pre-post design because evaluators repeatedly measure the same outcome for project participants, multiple times before and after project implementation. This approach will enable you to determine if trend lines change between the beginning and ending of project implementation.

For example, an Educational Technology, Media, and Materials (ETechM2) project delivered teacher professional development focused on a technology-based early literacy intervention for students with disabilities. One of its evaluation questions is, "Did professional development on use of technology in early literacy instruction affect student performance?" How might this project use a pre-post design to answer this question? The evaluation team might use an interrupted time series design to examine the effects of their project activities by collecting progress monitoring data on students' literacy acquisition monthly for 4 months before and monthly for 4 months after teachers participated in the professional development activities. The trend lines before and after the professional development might be compared visually and using regression analyses. The results would indicate whether students' literacy scores increased more sharply after the professional development, which would suggest that it was successful.

PRE-POST DESIGNS: RIGOR AND RESOURCES

(Rigor vs. Resources to Implement)
- One group post-test only
- Nonequivalent post-only control group
- One-group pre post-design
- Nonequivalent pre-post control group
- Interrupted time series
- Interrupted time series with a nonequivalent control group
- Randomized controlled trial

DROPOUT PREVENTION PROGRAM EXAMPLE: INTERRUPTED TIME SERIES

The Technical Assistance and Dissemination Center on dropout prevention from the first example could also make its pre-post design more rigorous by using an interrupted time series design to answer the same evaluation question: "Did intensive technical assistance on dropout prevention improve student attendance and academic achievement in participating districts?"

To use an interrupted time series design, evaluators could collect student attendance and progress monitoring data at least monthly for 5 months before and after the intensive technical assistance is delivered. This way, instead of comparing only one pre-score and one post-score, evaluators could compare five pre-scores and five post-scores. Regression analyses could be used to identify statistically significant changes in the slope of trend lines from pretest to posttest. The results from the interrupted time series design would strengthen the argument that changes were attributable to the dropout prevention program (compared to the strength of the argument that could be made with the single group pre-post design). To even further strengthen the interrupted time series design, the evaluators could collect the same 10 months of data for a nonequivalent control group of districts or schools that did not receive intensive technical assistance. As with the previous examples, evaluators could still follow up the data collection with qualitative interviews. However, because the nonequivalent time series design is more rigorous, the evaluators could use these interviews to determine which program components the participants thought were most impactful to inform replication.

TOPIC 2: COMPARED TO WHAT? EXTANT DATA

Although incorporating pre-post designs into your evaluation plan can be a great way to produce comparison data, it may not always be feasible to collect the pre-post data yourself. Fortunately, there are other potential sources of comparison data that you might use. Federal agencies, states, districts, schools, and early intervention programs collect, maintain, and will often share program data that might be useful for your evaluation. Data from these sources may even be available for your project participants (e.g., teachers) on an ongoing basis to inform practice moving ahead. This data is called "extant" data because it already exists and you do not need to collect it yourself.

Extant data is often used to complement primary data collection efforts, adding an important source of information that can fill gaps when you cannot collect all of the data you need. For example, you may be able to access extant student outcome data and examine outcomes that would otherwise be infeasible due to financial constraints. You might use extant data for all of the pre-post designs described above, if you can identify appropriate datasets. Potential sources of extant data relevant to OSEP-funded programs and projects include the Individuals with Disabilities Education Act (IDEA) Section 618 data, the Common Core of Data, the National Assessment of Educational Progress, and state websites (Part B and Part C). For example, a State Personnel Development Grant (SPDG) project's long-term outcome was increasing the availability of certified special education teachers. One of its evaluation questions is, "Did the supply of certified special education teachers improve in jurisdictions that did or did not receive technical assistance?" How might this project use extant data to answer this question? The evaluators might use administrative data collected by the state for evaluation or to report to OSEP as part of the IDEA Section 618 data collections. This data could be used to compare the percentages of certified teachers in participating jurisdictions before and after those jurisdictions received intensive technical assistance and training, as well as in jurisdictions that did not receive such support.

ADDITIONAL CONSIDERATIONS

Pre-post designs and extant data can both be great ways to produce comparison data to inform your evaluation. However, there are several important considerations when adding comparison data: planning early, assessing fidelity of implementation, and ensuring data security.

Planning Early

Being well-organized is particularly important if you plan to use extant data or a pre-post design. Pre-post designs require careful planning because data collection windows are often time specific. In many cases, post-test data should also be collected within a certain window (e.g., after a certain amount of time relative to the pre-test or the intervention). For example, if you do not collect pre-test data before your project activities begin, you will lose your opportunity to have a clean pre-post design. If you would like your evaluation to include a comparison group, you will need to identify and recruit that group before you begin the intervention. Similarly, if you are planning to use extant data, you will need to locate, download, and carefully examine the files of interest as early as possible, to ensure that they include the data you need. You would not want to plan on using certain data as part of your evaluation only to find out later that the data is not in a format you can use or at the level you need (e.g., district vs. state).

Collecting Fidelity of Implementation Data

Implementation fidelity refers to the degree to which your project was delivered as intended. Collecting data on fidelity of implementation can help you determine whether your project activities were carried out as you had planned, which gives you insight into the results of your evaluation activities. For example, if you are using comparison data and you see (or don't see) changes between groups, you will want to know if the outcomes are related to your project activities as implemented and not just as planned.

Ensuring Data Security

If you plan to collect pre-post data or use sensitive (or nonpublicly available) extant data, you may need to develop processes to secure the data and protect confidentiality. Some suggested steps include:

• Using identification codes in place of identifying information (e.g., a random number instead of a participant's first and last name);

• Creating separate files for data and participant-identifying information;

• Using secure data transfer protocols; and

• Protecting and restricting access to stored electronic data.

Limitations

With pre-post designs, you may see changes that are not actually attributable to your project. This possibility can be assessed by carefully considering threats to internal validity: considering whether the observed results are due to your project or to other factors that you did not include. In each of the three briefs on pre-post designs, we highlight the most relevant threats to internal validity.

Most of the limitations of extant datasets are related to the fact that you did not collect the data yourself. Although relieving yourself of this responsibility may be beneficial in many ways, remember that you have no control over the quality of the data collection, entry, or validation processes. In addition, data may be missing or incomplete, either because of errors during data collection or because data was suppressed due to small cell sizes.

ADDITIONAL RESOURCES

For more information on planning your evaluation, selecting an evaluation design, collecting and analyzing fidelity of implementation data, and maintaining data security, please see the **Evaluating Special Education Programs: Resource Toolkit.** Other resources related to the evaluation of special education programs are available on the **OSEP IDEAs That Work** website. To learn more about identifying good comparison data, you may want to refer to the other briefs in this series: "Using One-Group Pre-Post Designs," "Using Nonequivalent Pre-Post Control Group Designs," "Using Single-Case Interrupted Time Series Designs," and "Using Extant Data."