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

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.

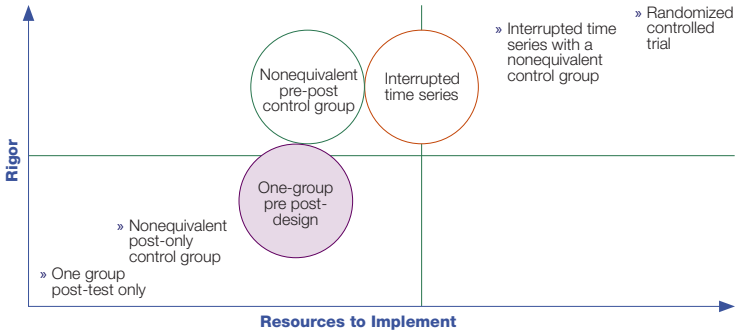
[1] In this case, an intervention refers to a product, service, or education program.

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

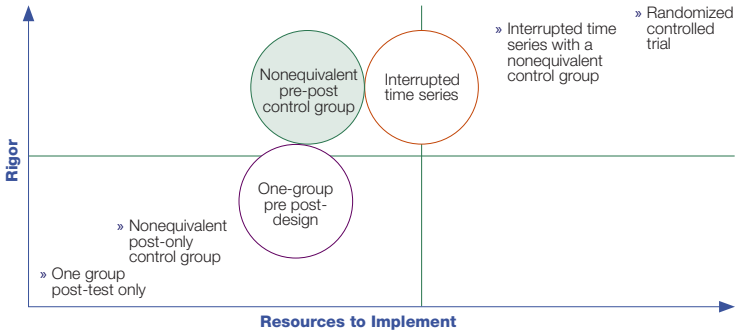
An OSEP Technical Assistance and Dissemination Center plans to provide intensive technical assistance on a dropout prevention program for students with disabilities. One of its evaluation questions is, "Did intensive technical assistance on dropout prevention improve student attendance and academic achievement 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.

PRE-POST DESIGNS: RIGOR AND RESOURCES



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

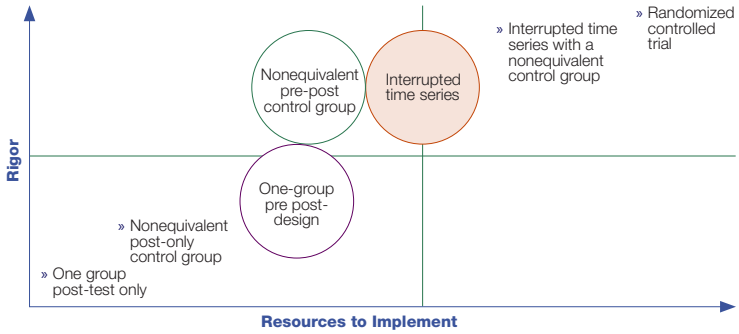
The OSEP-funded Technical Assistance and Dissemination Center on dropout prevention described in the previous example could make its pre-post design more rigorous by using a nonequivalent pre-post control group design to answer the same evaluation question: "Did intensive technical assistance on dropout 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.

PRE-POST DESIGNS: RIGOR AND RESOURCES



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.

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."

The Center to Improve Program and Project Performance has been funded with Federal funds from the U.S. Department of Education, Office of Special Education Programs, under contract number ED-ESE15-A-0016/0004. The Project Officers are Dr. Kristen Rhoads and Dr. David Emenheiser. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Compared to What?
Using Extant Data

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 would have been different if your project had not been implemented. That is, you need to know, "compared to what?" This brief focuses on using extant data to assess project results and represents the fifth brief in a five-part series, "Compared to What? Identifying Good Comparison Data to Assess Project Results."

Using extant data can make your evaluation more affordable and more rigorous. Federal agencies, states, districts, schools, and early intervention programs collect, maintain, and will often share data that may be useful for your evaluation. We call this extant data because it already exists and need not be collected by the end user. Knowing what relevant extant data is available can save your evaluation time and money, and you may be able to make comparisons that would not otherwise be possible because of resource constraints.

Extant data can complement primary data collection efforts by filling gaps when you can't collect all of the data you need or want. Used in this way, extant data can support rigorous evaluation designs. Extant data may allow you to:

- Examine outcomes that would not otherwise be possible to study. Extant data from early intervention providers, schools, districts, and states can all be beneficial, whether you are interested in student achievement, educational settings, teacher qualifications, or demographics. Primary collection of these types of data may often be difficult because of cost, timing, and necessary permissions for data collection. However, publicly available resources offer this type of extant data and can provide a wide selection of potential comparison groups, thus reducing logistical issues for your evaluation and minimizing burden to the child, parent, teacher, or school.
- Save money. Because extant data has already been collected as well as validated and organized into a usable format, using it can be less expensive than primary data collection. Therefore, you might be able to collect your own primary data for some outcome measures and complement it with other outcome measures that you assess with extant data.

IDENTIFYING SOURCES OF EXTANT DATA

There are many sources of extant data, and below we outline several that are likely to be relevant to programs and projects funded by the Office of Special Education Programs (OSEP). There are certainly additional sources, some of which you may discover while working in a early intervention program school, district, or state—so be on the lookout! Sources detailed on the following pages (complete with the types of data available and links) include data collected under the Individuals with Disabilities Education Act (IDEA), Common Core of Data (CCD), National Assessment of Educational Progress data, and state-level early intervention and special education data. The following section provides more detail about each data source.

IDEA Section 618 Data

A large amount of state-level data is collected under IDEA Section 618 and is publicly available on **OSEP's website**. This data, which the states collect and submit to OSEP annually, could be useful if your project anticipates observing changes at the state level. Longitudinal data on dropout/graduation, disciplinary removals, and teacher quality (among other topics) is available in state-level CSV files that go back as far as 2005, as well as in static data tables. This data is publicly available on the **IDEA Section 618 Data Products page** of the OSEP website. In addition, the IDEA Data Center (IDC) has an **Interactive Public Reporting Engine** that you can use to create easy-to-read charts and graphs from the IDEA Section 618 data. These charts and graphs may be useful as you explore your initial ideas, although you will have to download the files from OSEP to run any statistical analyses.

IDEA SECTION 618 DATA: IN-DEPTH

Part B Data include:

- Child Count: The number of children being served through IDEA Part B, by gender and disability.
- Educational Environments: The number of children aged 3-5 and 6-21, by educational environment.
- Personnel: The number of fully certified and not fully certified special education teachers and the number of related service providers fully certified and not fully certified, by specialty.
- Exiting: The number of students exiting special education and the reason for exit, by age, gender, race/ethnicity, English learner status, and disability.
- Discipline: The number of discipline events (removal, suspension, unilateral removal), by type of disability, gender, race/ethnicity, and limited English proficiency (LEP) status.
- Assessment: The number of students who took the regular assessment, with and without accommodations, in reading and in math; the number who took an alternate reading assessment, by alternate standard, grade-level standard; number of LEP students using English language proficient and modified standards; the number who took an alternate math assessment, by alternate standard, grade-level standard, and modified standard; the number not assessed in reading and math because of medical exemptions; and nonparticipants.
- Dispute Resolution: The number of complaints, mediation requests, due process complaints, written decisions within the timeline, written decisions within an extended timeline, expedited due process complaints, written settlement agreements, and expedited hearings.

Part C Data include:

- Child Count: The number of children being served through IDEA Part C, by race/ethnicity, gender, and age.
- Settings: The number of participating children, children in community-based settings, children in home-based settings, and children in other settings.
- Exiting: The reason for exit from early intervention, by race/ethnicity and gender.
- Dispute Resolution: The number of written, signed complaints; mediation requests; due process complaints; written decisions within the timeline; written decisions within an extended timeline; expedited due process complaints; written settlement agreements; and expedited hearings.

EXAMPLE USING IDEA 618 DATA

A Stepping Up Technology Implementation project developed a web-based toolkit with resources and strategies designed to help special education teachers respond to challenging behaviors in the classroom, with the ultimate goal of reducing the number of discipline events for special education students across school districts. The project staff worked to implement the toolkit in several districts in one state. They collected data in these districts on the number of discipline events by type of disability at the beginning of the school year, prior to the implementation of the web-based toolkit, and then at the end of the school year, after a full school year of implementation. Project staff analyzed the data for statistically significant decreases. In addition, the project staff used Section 618 discipline data at the state level to compare the district trends to statewide trends.

Common Core of Data

The CCD is a comprehensive, national database, updated annually, that provides basic information on all public elementary and secondary schools and school districts. The CCD database is the U.S. Department of Education's primary database on public elementary and secondary education in the United States. Fiscal and nonfiscal files are available in Excel, SAS, and flat text files and go back as far as the 1986-1987 school year. Data in these files includes a general description of schools and school districts, data on students and staff, and fiscal data. Housed the Institute of Education Science, the data are **available for download**.

CCD DATA: IN-DEPTH

Universal files provide information about the universe of states, school districts/local education agencies (LEAs), and schools, including data on staff counts and student enrollment. These are sometimes referred to as the universe files and are frequently used to draw samples.

- State-level data includes student membership counts disaggregated by grade, race/ethnicity, and gender, as well as staff full-time equivalent (FTE) counts by professional category.
- School district/LEA-level data includes student membership disaggregated by grade, race/ethnicity, and gender; staff FTE counts by professional category; and counts of English learners and children with disabilities.
- School-level data includes student membership disaggregated by grade, race/ethnicity, and gender; staff FTE counts; and counts of students eligible for free and reduced-price lunches.

Dropout and school completion data files are available at the state and school district/LEA levels, showing the numbers of dropouts, high school diploma recipients, and other high school completers.

- State-level dropout and completion data files contain the numbers of dropouts from each of grades 9-12 and the relevant event dropout rates, as well as the numbers of high school diploma recipients, the number of other high school completers, and the relevant Averaged Freshman Graduation Rate.
- School district/LEA dropout and completion data files contain the numbers of dropouts from each of grades 7-12, as well as the numbers of high school diploma recipients and other high school completers.

EXAMPLE USING CCD DATA

An Educational Technology, Media, and Materials for Individuals with Disabilities (ETechM2) Center compared third grade reading outcomes for students who engaged with accessible digital instruction materials for a full school year, compared to students in schools or classrooms not exposed to the technology. The evaluators used CCD data to select a nontreatment comparison group that was comparable to the treatment group. To ensure that the groups were comparable, the evaluation team matched them on key characteristics in the CCD, including school size and student demographic information on type of disability, gender, and grade/age, as well as whether English was their primary language. Reading outcomes were then examined for both groups, and statistical analyses were used to identify any significant differences between the treatment and comparison groups.

National Assessment of Educational Progress (NAEP) Data

NAEP is the largest nationally representative and continuing assessment of students in the United States and is administered to a sample of students every year. NAEP data includes the test results for math, reading, science, writing, technology and engineering literacy, arts, civics, geography, economics, and U.S. history assessments; grades 4, 8, and 12; jurisdictions (states and selected districts); and student performance in the context of gender, type of school, location, race/ethnicity, eligibility for free or reduced-price lunch, students with disabilities, and students identified as English learners.
The data is housed at NCES.

NAEP DATA: IN-DEPTH

NAEP assessment results are presented using average scale scores and NAEP achievement levels.

- Average scale scores represent how students performed on each assessment. Scores are aggregated and reported at the student group level for the nation, states, and districts. They can also be used for comparisons among states, districts, and student groups.
- NAEP achievement levels are performance standards that describe what students should know and be able to do. Results are reported as percentages of students performing at or above three achievement levels (NAEP Basic, NAEP Proficient, and NAEP Advanced). Students performing at or above the NAEP Proficient level demonstrate solid academic performance and competency over challenging subject matter. Note that the NAEP Proficient achievement level does not represent grade-level proficiency as determined by other assessment standards (e.g., state or district assessments).

The actual data files are not publicly available for download, but the NAEP Data Explorer creates tables and graphics that you can customize to examine different sets of results.

- The **NAEP Data Explorer** (NDE) enables you to see the results of specific assessments (mathematics, reading, science, writing, technology and engineering literacy, arts, civics, geography, economics, and U.S. history) across multiple years and broken down by a variety of student groups.
- The main national and state results are available in those 10 subject areas going back to 1990.
- Four subjects—mathematics, reading, science, and writing—are assessed most frequently and reported at the state and district levels, usually for grades 4 and 8.
- For in-depth exploration, the NDE provides statistical results such as significance testing, gap analysis, and regression analysis. Users can export tables and charts to Word documents, Excel workbooks, and PDFs.

EXAMPLE USING NAEP DATA

A Technical Assistance and Dissemination (TA&D) Center delivered intensive math technical assistance in several NAEP Trial Urban District Assessment (TUDA) districts. The project staff wanted to investigate whether math proficiency increased for students with disabilities in the districts that received the technical assistance for 2 years compared to students in other TUDA districts throughout the nation. The most recent NAEP math scores as well as achievement levels for students with disabilities in grades 4 and 8 in the intervention districts were compared with NAEP scores from other TUDA districts. In addition, the project looked at the change in scores and proficiency levels in the treatment and comparison groups over the last 8 years, giving the project team four data points since the NAEP data is collected every other year.

State Early Intervention and Special Education Data

State early intervention and special education websites can also be a rich source of extant data. The data available will differ by state; if you are working with a specific state, it will always be worthwhile to examine its website for relevant, publicly available data that you might be able to incorporate into your evaluation. To provide a sense of the types of unique data available across different states, we highlight below some of the publicly available data in Florida, Massachusetts, South Carolina, and Nebraska. In addition, districts and schools usually collect their own data such as progress monitoring data collected regularly on all students. However, this data is not likely to be publicly available; to access it, you would probably need to apply to the state or district and complete a data usage agreement.

PUBLICALLY AVAILABLE STATE-LEVEL DATA: IN-DEPTH

- **LEA profiles.** These profiles often include demographic and district-level data that mirrors the state-level IDEA Section 618 data discussed above. For example, Florida, Massachusetts, South Carolina, and Nebraska's Early Development Network all have a version of an LEA or regional profile. Massachusetts also posts district- and school-level data on teacher salaries, per-pupil expenditures, and educator evaluation.
- **Statewide assessment data.** Some states make their statewide assessment data publicly available. In Massachusetts, aggregated data is presented for all students, while in other states, like Florida and South Carolina, data for students with and without disabilities is presented separately. For example, Florida has Data Books for each school year (dating back to 2008) that contain aggregated assessment data for students with disabilities and data at the district level for grades 3-10 reading (disaggregated by grade) and grades 3-8 math (disaggregated by grade). South Carolina's Department of Education makes available math and reading assessment data for students with Individualized Education Programs (IEPs) by school in the following categories: percentage below basic, percentage basic, percentage proficient, percentage advanced, and percentage proficient/advanced. This data is presented in Excel files and contains percentages for all students, as well as for students with disabilities who take the regular assessment (with and without accommodations) and alternate assessments. This data is reported for each school year, beginning in 2011-2012.
- **Early intervention data.** Early Intervention (i.e., Part C) data is also publicly available on state websites; however, depending on where the department is housed, some data may be included on the Department of Education's website or on the website for the Early Intervention Program. For example, Florida Part C early child outcomes data is presented on Florida's Exceptional Student Education website, which provides data on the percentage of children who entered PreK below expectations but substantially increased their rate of growth in PreK, as well as the percentage functioning within age expectations by the time of PreK exit (by district and developmental domain). In contrast, Nebraska data is available on the state's Early Development Network's website, which has regional data beginning in 2005-2006 and includes each region's Part C Annual Performance Report.

EXAMPLE USING STATE SPECIAL EDUCATION DATA

A Personnel Development Program (PDP) project implemented a new induction and mentoring program to increase retention of special education teachers who work in high-need schools. The evaluators used extant state and district databases to determine the percentage of teachers who participated in the induction and mentoring program that were retained in their high-needs schools after 2 years, and how long they remained in those or other high-needs schools (i.e., up to 5 years).

EXAMPLE USING LOCAL DATA

A Personnel Development Program (PDP) project uses shared coursework, group assignments, and coordinated field experiences to prepare scholars in the special education and school psychology degree programs to use evidence-based practices in serving children with disabilities. The project staff wants to evaluate the effect of their interdisciplinary approach in building collaboration skills. The evaluators used local extant data from a nearby large urban district that took on many of the university's interns. Field supervisors in the district used a locally created rubric to assess their interns' interdisciplinary collaboration skills at the completion of their internship. The evaluators compared the percentage of proficient project-funded interns (i.e., 3.0 on a 5.0 scale) who completed the coordinated field experiences, to interns from the same degree program, placed in the same district, who did not participate in the interdisciplinary PDP project.

INTEGRATING EXTANT DATA INTO YOUR EVALUATION DESIGN

Determine how you want to use extant data in your evaluation. It may be that relevant outcome data on your population is already being collected, and it will be sufficient for your evaluation. Not only will using extant data potentially save time and money, but it may allow you to look at a larger sample than would be possible if you were collecting the data. Additionally, you may be able to use extant data to create comparison groups of those not participating in (or not expected to be affected by) your program or project. In fact, the use of extant data may facilitate using a comparison group because collecting original data from demographically similar, nonparticipating students, teachers, schools, districts, or states can be particularly challenging.

Select an appropriate extant dataset. Once you have decided how you plan to use extant data, you will need to select an appropriate dataset. There are several factors to consider:

- Is the extant outcome data relevant to your evaluation?
- Is the extant data available over a time period that will work for your evaluation?
- Do the extant files use a unit of analysis appropriate to answer your evaluation questions?
- Are the extant data files in a format with which you can work?

Plan for and conduct your analyses. Analyses of extant data can be relatively straightforward, or they can be quite complicated. We cannot detail all of the analytic options in this brief because the choice depends on your dataset, outcomes of interest, and evaluation questions. To help you decide how best to proceed, we suggest working with a statistician who has the relevant expertise. Regardless of the statistical approach you use, you'll first want to dig into your extant data file and identify the variables associated with your desired outcomes. A codebook associated with the data file may help you better understand each variable (e.g., definition, range of potential values).

If you are using extant data to create a nontreatment comparison group, you will want to use an approach called matching. Matching can help you improve the similarity of the treatment and comparison groups, thus increasing the likelihood that any differences you find might be due to your intervention. Matching involves identifying important variables that you think might contribute to differences in your outcome and grouping participants so that the treatment and comparison groups have very similar characteristics on these variables. You'll want to use matching variables that are stable and reliable and that are correlated with the outcome variable. You can then compare the comparison and treatment units on the outcome variables you have identified. We recommend that you also involve a statistician with this process.

CONSIDERING LIMITATIONS

In general, the biggest advantage of extant data is also its biggest limitation – you did not collect the data yourself. Because you did not collect the data, it can be difficult to achieve a perfect match between your project's anticipated outcomes and the outcome data that is available in extant datasets. In addition, data may be missing or incomplete, either because of errors during data collection or because data has been suppressed due to small cell sizes. Finally, you don't have any control over the format of the data, and sometimes the way that the data is organized will limit the types of analyses you are able to conduct. Therefore you must plan early and consult with a statistician as needed if you decide to use extant data as part of your evaluation. As early as possible, locate, download, and carefully examine the files in which you are interested. You will not want to plan on using certain data as part of your evaluation only to find out later that it is not usuable.

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 Nonequivalent Pre-Post Control-Group Designs," and "Using Single-Case Interrupted Time Series Designs."

The Center to Improve Program and Project Performance has been funded with Federal funds from the U.S. Department of Education, Office of Special Education Programs, under contract number ED-ESE15-A-0016/0004. The Project Officers are Dr. Kristen Rhodes and Dr. David Emenheiser. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

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.

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.

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.

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."

The Center to Improve Program and Project Performance has been funded with Federal funds from the U.S. Department of Education, Office of Special Education Programs, under contract number ED-ESE15-A-0016/0004. The Project Officers are Dr. Kristen Rhoads and Dr. David Emenheiser. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Compared to What?
Using One-Group Pre-Post Designs

Measuring outcomes for the population affected by your project is important, 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 one-group pre-post designs and represents the second 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 complete some type of assessment before the project is implemented and afterwards. Measures can 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 pre-test to 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.

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-test.

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 while a one-group pre-post design is less rigorous than some other pre-post designs, it does not require complex statistics. In the one-group pre-post design, the group participating in your project (or the group expected to be affected by your project) is measured before the project is implemented and then afterwards, so there is a point of comparison that allows you to assess changes (such as changes in knowledge or behavior). This design is preferable to and more rigorous than a post-test-only design because, with information about the starting point, you can identify changes that occurred and attribute them to your treatment with greater validity. However, if you are using a one-group pre-post design and you see changes from pre- to post-test, there may be other reasons for the changes.

One way to improve a one-group pre-post design is to add a nonequivalent dependent variable. This means that, in addition to your pre- and post-test, you could measure the change in a similar variable relevant to your population that you would not expect to change as a result of your project. For example, if students' math scores rose at the same rate as reading scores, when your project intervention was focused solely on reading, something other than your project, such as maturation or testing effects, might be causing the reading score increases. If, on the other hand, only students' reading scores increased, you could be more confident that your project intervention was the likely reason for the change in reading scores.

AN EXAMPLE OF A ONE-GROUP PRE-POST DESIGN

A Parent Training and Information (PTI) center worked throughout a state to increase its outreach to parents and families who live in rural areas. After 6 months of intensive outreach, the PTI analyzed the demographic data of families to whom they provided technical assistance and compared it to the demographic data from the previous 6 months to determine if they were reaching an increased number of families in rural areas of the state. To enhance the internal validity of its study, the PTI added a nonequivalent dependent variable by examining the change in parents and families served from other suburban and urban areas in their state. The PTI had greater confidence in the success of its outreach intervention after finding no change to the numbers of families served in other geographical areas but large increases in numbers of families served in rural areas of the state.

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 Nonequivalent Pre-Post Control-Group Designs," "Using Single-Case Interrupted Time Series Designs," and "Using Extant Data."

The Center to Improve Program and Project Performance has been funded with Federal funds from the U.S. Department of Education, Office of Special Education Programs, under contract number ED-ESE15-A-0016/0004. The Project Officers are Dr. Kristen Rhoads and Dr. David Emenheiser. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

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.

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.

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.

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."

The Center to Improve Program and Project Performance has been funded with Federal funds from the U.S. Department of Education, Office of Special Education Programs, under contract number ED-ESE15-A-0016/0004. The Project Officers are Dr. Kristen Rhoads and Dr. David Emenheiser. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.