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.

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."