CIPP

Demonstrating Evidence

Across the Project Cycle

Jill Lammert
Sarah Heinemeier
Bethany Howell
Amy Germuth
Tom Fiore

Westat®

About This Tool

This Tool was developed as part of the Center to Improve Project Performance (CIPP) operated by Westat for the U.S. Department of Education, Office of Special Education Programs (OSEP). The authors thank the OSEP, Westat, and other individuals who provided input.

Suggested Citation

Overview of the Center to Improve Project Performance

First formed in 2008, CIPP's overall mission is to advance the rigor and objectivity of evaluations conducted by or for U.S. Department of Education's Office of Special Education Programs (OSEP)-funded projects so that the results of these evaluations can be used by projects to improve their performance and used by OSEP for future funding decisions, strategic planning, and program performance measurement.

CIPP provides evaluation support, oversight, and technical assistance (TA) to OSEP projects. CIPP staff work with project and OSEP staff to refine project logic models and develop evaluations. Based on the evaluation design and plan, CIPP staff have overseen evaluation activities and provided grantee technical assistance (TA), as needed, including selecting samples; developing draft instruments; assisting in new instrument pilots, identifying appropriate data collection and analysis strategies; performing reliability checks; providing accurate descriptions of the methods and valid interpretations of findings; and organizing, reviewing, and editing project evaluation reports. In addition to providing TA to OSEP-funded projects on request, CIPP staff prepare a variety of TA products focused on evaluation issues, and deliver presentations on evaluation through webinars and conferences.

Contact Information

https://www.cippsite.org/SitePagesPublic/Home.aspx
1-888-843-4101

Thomas Fiore, CIPP Project Director
Westat
ThomasFiore@westat.com

Jill Lammert, CIPP Deputy Project Director
Westat
JillLammert@westat.com

Patricia Gonzalez, Project Officer
Office of Special Education Programs
U.S. Department of Education
Patricia.Gonzalez@ed.gov

IDEAs that Work
Office of Special Education Programs
U.S. Department of Education

Contents

Instructions: Click on a topic to navigate to that section. You can use the "Back" button in the upper left-hand corner of the page to navigate backward.

Tables

Figures

Boxes

Introduction

Increasingly, federal grantees and their evaluators are asked to demonstrate "evidence" that their projects and the interventions they promote are achieving their objectives. But evidence is a term that can be applied in varying ways. For example, when selecting an intervention to improve the reading skills of elementary-aged children, educators are often asked to identify an "evidence-based" or "research-based" practice. When applying for federal funding, applicants might need to show that the kindergarten intervention they intend to implement shows "moderate evidence of effectiveness" or "evidence of promise" at improving the behavior of children with disabilities in an inclusive setting. When evaluating a project, a grantee will need to provide evidence that they have achieved their desired outcomes. Schwandt defines evidence as "information bearing on whether a belief or proposition is true or false, valid or invalid, warranted or unsupported."[1] Given the myriad ways the term evidence can be used, you might be asking where you can find such evidence and how you will determine its quality. If no evidence yet exists, you might wonder how you can produce it. Or, if the project you are designing, implementing, or evaluating is targeted at changing adult behaviors, how can you show evidence that the project ultimately improves outcomes for children or students with disabilities?

This tool discusses the different types of evidence that can be used throughout the Planning, Implementation, and Results Phases of a project to show whether and to what degree you're achieving your project objectives. For example, the tool details how to use evidence to support theories of action, guide development of logic models and choice of interventions, and provide information on how the project activities are progressing. It also addresses how to use the evidence you collect to monitor progress in project implementation, provide formative feedback, and show results at a particular phase—with the goal of demonstrating ultimate project results, or long-term outcomes. Throughout, we'll address the following questions: What is the nature of the evidence needed and available at the different project phases? How can evidence be used to inform project management and evaluation?

Organization and Use of This Tool

The figure below illustrates one way evidence can be conceptualized across the different phases of the project cycle. As can be seen, the discovery and use of evidence is an iterative and somewhat cyclical process. Starting with the *Planning Phase*, you create a plan, supported by evidence, that clearly outlines the linkages between the proposed project activities, outputs, and expected short-term, intermediate, and long-term outcomes (Depending on the project, these outcomes could relate to adults, teachers, related-service providers, or to infants, toddlers, children, and youth with disabilities or their families.)

Discovery and Use of Evidence across the Project Cycle

DISCOVERY

USE

Planning

Implementation

Results

Instructions:
Click on a phase to navigate to that section.

---
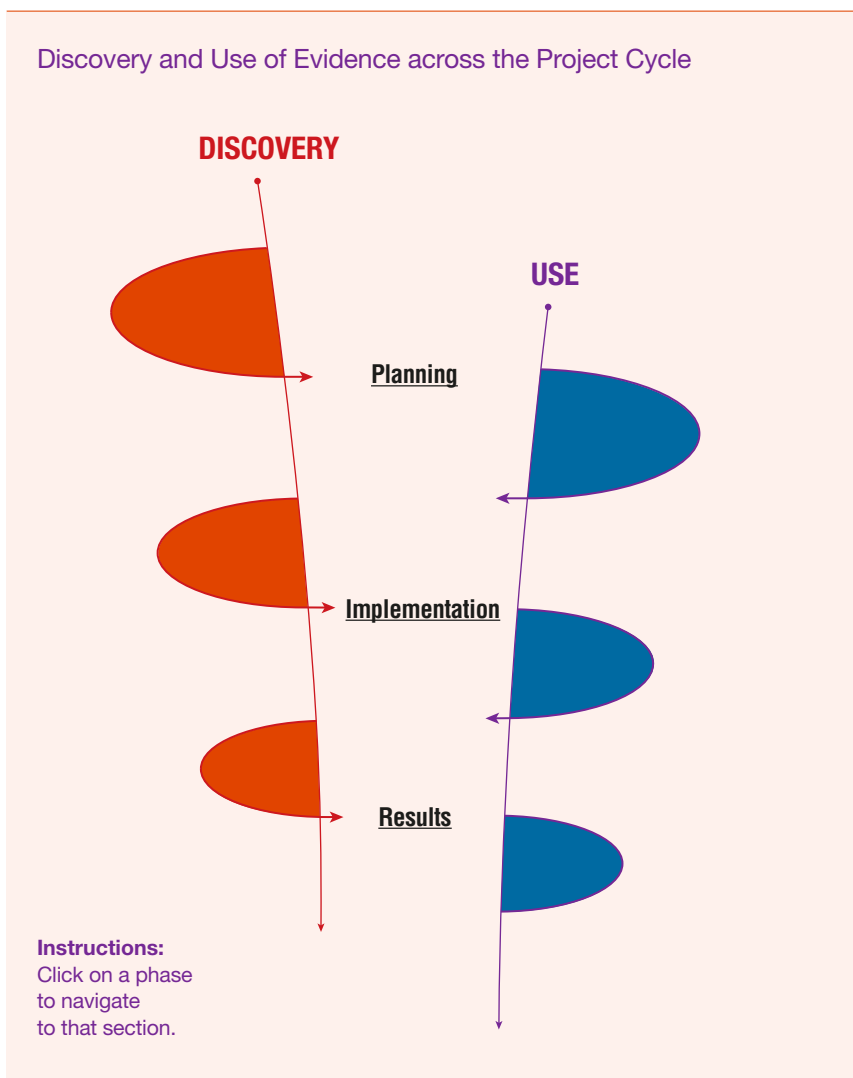
[1] Schwandt, T. A. (2009). Toward a practical theory of evidence for evaluation. In Stewart I. Donaldson, Christina A. Christie & Melvin M. Mark (Eds.) *What counts as credible evidence in applied research and evaluation practice?* Thousand Oaks, CA: Sage Publications, Inc.

As you plan how you'll implement and evaluate your project, you must think through the evidence you'll need at the different phases to provide feedback to support ongoing implementation and to demonstrate that you're achieving your goals. This will help you to develop a project plan—including a detailed evaluation plan—that can help guide you as the project progresses.

As you move to the *Implementation Phase*, you'll gather information to monitor progress and provide feedback on implementation of strategies, activities, and outputs. The information you gather during the Implementation Phase can help you improve the way you carry out the planned strategies and activities, and can give you early indications of your progress toward achieving the project's expected outcomes. Using the information you collect during this phase, you can make any needed adjustments to the project, so as to improve project functioning and increase the likelihood of achieving your outcomes.

Finally, the *Results Phase* focuses on evaluating project outcomes. As we'll discuss, many projects identify short-term, intermediate, and long-term outcomes.[2] During this phase, you'll be collecting data about the effectiveness of your strategies and activities in producing the expected outcomes. In many situations, you'll only be collecting data and evaluating achievement of short-term and intermediate outcomes. But, we'll talk through how you also can use data to demonstrate whether and how project activities, outputs, and short-term and intermediate outcomes are likely to result in achievement of long-term outcomes. It's important to note that these phases aren't entirely sequential, nor are they mutually exclusive. Evidence gathered—or produced—during one phase can be used to inform improvements in or plan for other phases.

Using This Tool

The figure on the previous page is designed as an interactive graphic that you can use to navigate this tool. You also can search for what you're looking for by reviewing the *Table of Contents (page 3)*. By clicking on the different project phases in the graphic you can get to brief discussions of specific topics related to discovery and use of evidence.

For example, a click on "Implementation" will take you to a discussion of the types of evidence that might be collected during the Implementation Phase (e.g., stakeholder input), and how to use that evidence to provide formative feedback or help assess fidelity during the Implementation Phase.

At the top of each section we have added links to the topics in that section so you can quickly navigate to a particular topic of interest. Throughout the document, boxes with **ORANGE** headers on the left side of the page, identify information and resources related to the discovery of evidence, while boxes with **BLUE** headers on the right side of the page, relate to the use of evidence. If you click on a link within the text, you can use the "Back" button found in the upper left-hand corner of the page to navigate back to the place where you started. Also, you can use the arrows found in the lower right-hand corner of the document to navigate to the Previous or Next page.

---

[2] These are also called direct, medium-term, and distal outcomes. We prefer the term "intermediate" outcomes instead of medium-term, so we will use that throughout.

Thinking about Evidence

"Evidence" can be thought of in multiple ways. For example, the term can apply to the information you might gather when you're thinking through a project's *theory of change* or starting to develop the project *logic model*—such as conducting a review of the theoretical literature or consulting practice guides to identify possible implementation strategies. It also can apply to the prior research that shows that a certain intervention or approach is effective (or not effective) in a given setting or with a particular group of individuals. Similarly, data collected during a project's implementation and formative evaluation serve as sources of evidence about the project's progress or outcomes, while summative evaluation data can demonstrate whether and to what extent a project has ultimately achieved its objectives.

Demonstrating "Evidence" for Education Projects and Interventions

➤ The **U.S. Education Department General Administrative Regulations (EDGAR)** outline *different levels of evidence* that applicants for federal funding might be required to use to support the selection and use of a particular educational intervention or approach in their applications. These levels differ according to the rigor and scope of the research evidence supporting project effectiveness.

➤ The 2016 **Every Student Succeeds Act (ESSA)** defines what constitutes an *"evidence-based" practice* in education contexts.

➤ The U.S. Department of Education's (Department) **What Works Clearinghouse (WWC)** serves as a source of scientific evidence for "what works" in education.* Using standards developed by experts in education and research methodology, the WWC conducts critical reviews of thousands of education studies. The WWC is a good source of evidence on educational interventions that have been formally studied using rigorous research designs. To date, out of the 11,323 studies that have been reviewed by the WWC, 1,935 have been screened using the WWC's "Children and Youth with Disabilities" review protocol. The great majority of those studies were not considered eligible to be reviewed by the WWC for various reasons. In fact, only 98 studies (5 percent) reviewed using that protocol were considered to feature a high-quality research design that could demonstrate any evidence of effectiveness.** For information on the *WWC evidence ratings*, see the Appendix.

➤ The National Science Foundation and U.S. Department of Education's Institute of Education Sciences (IES) developed **Common Guidelines for Education Research and Development** in which they outlined the ways various types of research might generate evidence about strategies and interventions for increasing student learning.

---

* Source: What Works Clearinghouse. (2015). About the WWC. Washington, DC: U.S. Department of Education, Institute of Education Science, What Works Clearinghouse. Retrieved from *http://ies.ed.gov/ncee/wwc/aboutus.aspx*

** Source: *http://ies.ed.gov/ncee/wwc/ReviewedStudies.aspx*

The framework for thinking about evidence outlined by Puddy & Wilkins[3] and illustrated in the figure to the right offers a way to look at the various types of evidence that are available to project staff and evaluators. This framework identifies three overlapping types, or facets, of evidence:

- **Best available research evidence**—Information derived from scientific inquiry. Examples include WWC reviews or data collected during a summative evaluation that show to what extent and how well a project has achieved its goals.

- **Contextual evidence**—Measurable factors in a specific context that may impact the success of a particular strategy or intervention, including whether a particular activity is likely to be acceptable, feasible, and useful in a local setting. Political climate, state or district regulations, or school-level factors such as high staff turnover are some examples of this type of evidence.

- **Experiential evidence**—Experience and expertise of those who have practiced or lived in a particular setting. Examples of this type of evidence include expert opinion on the appropriateness of an intervention for a particular target population and practitioner experience related to how similar interventions have been implemented in similar or different settings.

While clearly important, research evidence is just one source of evidence that project staff and evaluators may use to determine which interventions, approaches, or practices to use, or to assess to what degree a project is achieving the desired outcomes. Contextual and experiential evidence (which can be collected during the preparation of a project plan or during the project evaluation) can also be used to help project staff and evaluators to know whether a particular project or intervention is working—or is likely to work—in a particular context. Indeed, as Puddy and Wilkins point out, *"evidence-based decision making occurs when the best available research evidence is combined with the experiential evidence of field-based expertise and contextual evidence."*[4]

An important consideration as you think about how you will discover and use evidence for your project is that the "best available research evidence" for a given project, intervention, policy, or practice may not meet established standards of research rigor (such as those outlined by

Framework for Thinking about Evidence

Source: Puddy & Wilkins, 2011.
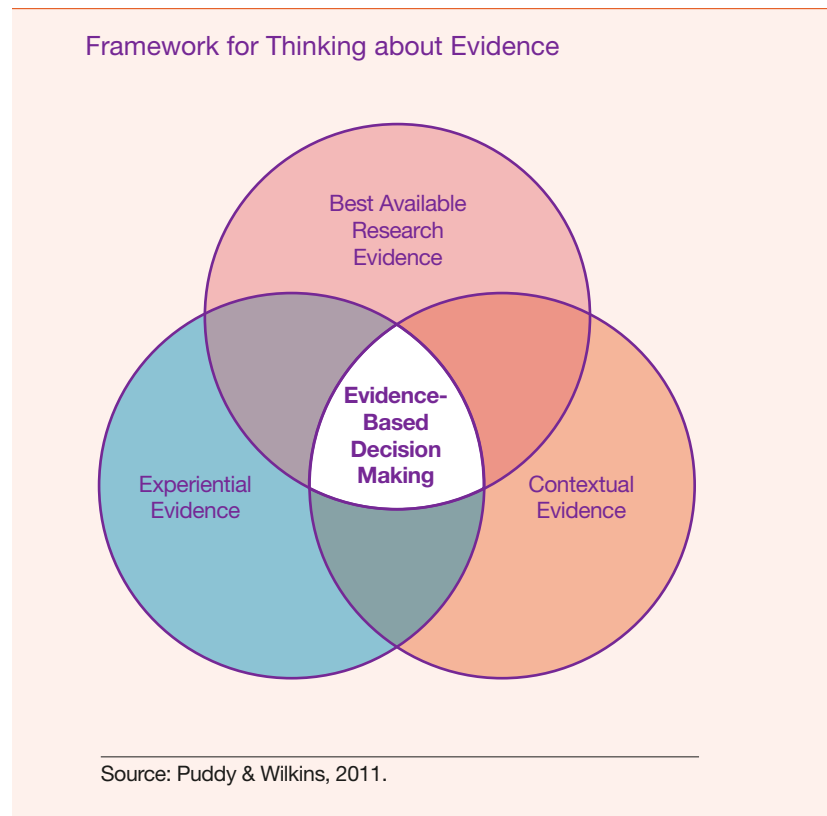
the WWC and the NSF/IES). Further, depending on the type of project being developed or implemented, all three types of evidence may not be available in all phases. For example, there may not be substantial prior research into a specific technical assistance (TA) approach being used by a project, so contextual and experiential evidence, and possibly some basic evaluation findings, might be the only sources of information for project staff choosing how to provide high-quality TA to parents of children with disabilities. On the other hand, there may be numerous studies examining in a controlled laboratory environment how a specific reading intervention works for children with disabilities, but there may be little contextual or experiential evidence showing how well that intervention might work in an inclusive general education classroom. In this document, we'll point out how you might discover and use these different types of evidence in the various phases of your project.

[3] Puddy, R.W., & Wilkins, N. (2011). Understanding evidence part 1: Best available research evidence. A Guide to the Continuum of evidence of effectiveness. Atlanta, GA: Centers for Disease Control and Prevention. Retrieved from *http://www.cdc.gov/violenceprevention/pdf/understanding_evidence-a.pdf*

[4] Puddy & Wilkins 2011, p. 4.

Planning Phase
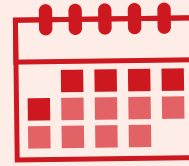
In this section, we'll briefly discuss how you can collect and use evidence to plan a project, including:

➤ Identifying needs the project will address

➤ Identifying available resources to support the project

➤ Developing or modifying a theory of change, including:

  • Selecting a project intervention

➤ Creating a logic model, including:

  • Selecting specific strategies or activities

  • Identifying realistic outputs and outcomes

➤ Incorporating evaluation evidence into the project

How do I collect and use evidence to plan a project?

In this section, we describe project planning as a set of activities that occur prior to project implementation. However, we also recognize that planning may be an iterative and recurring process as projects respond to unanticipated events and challenges as the project progresses. To begin, let's start with the end in mind. How does knowing what we want our end product to be help us with both the discovery and use of evidence? A good place to start is to develop a detailed plan that will serve as a guide for implementing a project. The plan is informed and supported by the available research, contextual, and experiential evidence and clearly outlines the linkages between proposed project activities, outputs, and expected short-term, intermediate, and long-term outcomes. The plan is a step-by-step guide to the "who, what, where, when, and how" of a project, tying together the approach to carrying out project activities and to collecting, managing, aggregating, analyzing, and reporting data on implementation and outcomes. A plan will likely include:

• the project rationale (i.e., data that demonstrate the needs the project is expected to address);

• a theory of change and logic model;

• evaluation questions and activities (including data collection and analysis);

• performance measures and implementation guidelines; and

• a comprehensive timeline of all activities, including evaluation and performance monitoring.

As you identify a theory of change and develop a logic model, you should use research, contextual, and experiential evidence to identify needs and support your choice of project objectives, intervention, strategies and activities, outputs, and outcomes. In doing so, you're well along the way to putting together your project plan. In fact, when done well, the logic model can guide you to develop a comprehensive project plan that outlines in detail how you will implement the various strategies and activities to produce the expected outputs and outcomes.

Discovery and Use of Evidence in the Planning Phase

What Do I Need?	How Might I Use It?
➤ Evidence of needs and resources to support the project plan ➤ Evidence to support the project's underlying theory of change and guide development of the logic model ➤ Evidence to select a particular intervention or combination of interventions ➤ Evidence to support use of particular strategies or activities for implementing the project	➤ Match the project plan to identified needs ➤ Refine or confirm the project's underlying theory of change ➤ Create a high-quality logic model for project planning and evaluation—make links between planned activities, strategies, and shorter-term and longer-term outcomes

How do I identify the needs the project will address?

Part of the process of discovering evidence for project planning is finding data the project team can use to develop specific statements about needs. In planning a project, it's important to understand and document the following:

- The level or scope of the need in the community or target population of interest.

- Upward or downward trends in the need.

- Why the need exists in the community or target population of interest.

- If and how needs vary among sub-groups of the community or target population.

- The context in which the project will be implemented.

The resources described below may serve as sources of evidence you can use to identify needs, such as population counts, measures of academic achievement, estimates of the prevalence of different special needs groups with the population, etc. When available, you might also use existing needs assessments and studies, such as:

- Relatively current (i.e., within the last three years) needs assessments completed for the community or target population.

- Relatively current gap analyses, which establish the missing links between needs and community resources.

- Policy briefs, issued by elected representatives, policy think tanks, professional organizations, or lobbyists that can provide information on the project's policy implications and the general level of public support for such a project.

Selected Resources to Help Identify Needs

➤ **U.S. Department of Education Resources**

- ✦ Technical Assistance Centers *https://www.osepideasthatwork.org/find-center-or-grant/find-a-center*

  - ✦ Comprehensive Centers *http://www2.ed.gov/about/contacts/gen/othersites/compcenters.html*

  - ✦ Regional Educational Laboratories *https://ies.ed.gov/ncee/edlabs/regions/*

  - ✦ Rehabilitation Research and Training Center on Disability Statistics and Demographics *http://disabilitycompendium.org/statistics*

  - ✦ National Center on Educational Outcomes *https://nceo.info/*

  - ✦ National Center on Intensive Intervention *http://www.intensiveintervention.org/*

- ✦ OSEP GRADS 360 *https://osep.grads360.org/#program*

- ✦ National Center for Education Statistics *http://nces.ed.gov/*

  - ✦ Common Core of Data *https://nces.ed.gov/ccd/*

  - ✦ Early Childhood Longitudinal Study *https://nces.ed.gov/ecls/*

  - ✦ National Assessment of Educational Progress *http://nces.ed.gov/nationsreportcard/about/*

  - ✦ The Condition of Education *http://nces.ed.gov/programs/coe/*

  - ✦ The Digest of Education Statistics *https://nces.ed.gov/programs/digest/*

➤ **State and District Education Databases**

  Look for those that are publicly available and updated regularly (e.g., annually). Consider contacting state and district data liaisons for more information on data that might not be publicly available.

➤ **Professional Organizations**

- ✦ National Association of Special Education Teachers (see, for example, the Resources page *https://www.naset.org/786.0.html*

- ✦ Council for Exceptional Children *https://www.cec.sped.org/*

- ✦ National Center for Learning Disabilities *http://www.ncld.org/*

➤ **Publications, Policy Statements, and Guidance Documents**

- ✦ Special reports such as Racial and Ethnic Disparities in Special Education

A Note on Data Quality

When thinking about the quality of data (whether it be existing data you use to identify needs or data you produce through your project), you should look for these five characteristics of high-quality data and ask yourself the following questions:

➤ **Precision.** To what extent do the data collected reflect an exact measurement (for quantitative data) or include the narrative information (for qualitative data) needed to respond to a question?

➤ **Accuracy.** To what extent do data reflect the actual value of an observation or achievement? For example, is a measure of height or weight accurate or "off" by several inches or pounds? Do the data provide a "true" (e.g., verifiable) account of a phenomena or experience?

➤ **Reliability.** To what extent can a data measurement be replicated with accuracy and precision? When gathering data from individuals, might a person have any reason to respond falsely?

➤ **Consistency.** To what extent do data or individuals' responses agree with each other?

➤ **Completeness.** To what extent is complete information provided? For example, is the unit of measurement provided to help interpret the data? Is there enough context to understand an individual's response?

If a given resource does not provide any information that allows you to judge these aspects of the data, proceed with caution—it might not be a trustworthy source. The same goes for data produced through a project evaluation: If any of these data quality characteristics are lacking, be careful about the conclusions you draw from the data.

What resources might be available to support the project?

In the Planning Phase, it is also important to gather evidence and data on the resources, or "inputs," that will be available throughout the life of the project. Inputs may include the following types of resources and materials:

• Funding

• Internal resources (e.g., staff, time, facilities)

• External resources (e.g., expertise, technical assistance)

• Community resources and supports (e.g., volunteers)

• Guidelines and protocols for project implementation, such as implementation guides, recommended or required assessments, and technical support staff

To do this, you can search for research and contextual evidence related to past projects (e.g., evaluation reports, conference presentations) to identify resources that were used to support project functioning, or you can gather experiential evidence by consulting key stakeholders to learn what types of resources may be needed and available for your project. Project Officers or funders might also be good sources of information for contextual and experiential evidence related to project resources.

How do I identify a theory of change for the project?

Another way evidence can be used during planning is to identify a theory of change for the project. A **theory of change** outlines the goals and objectives of a project as well as the various processes and contextual influences that are expected to result in outcome achievement. While you may not have a formal theory of change for your project, you undoubtedly have an idea of how you think your project will work to address the identified needs. The key is to document the mechanisms by which the various actors, change processes, and external influences are expected to combine to produce the expected results.

Some interventions or programs already have an established research base and theory of change that can be used as evidence for the selection of your theory of change. For example, the Positive Behavioral Interventions and Supports (PBIS) framework provides implementation *blueprints* for districts and schools. However, for many projects there may be little research evidence related to effective programs or strategies for working with a particular group, the factors that affect a project's ability to deliver services, or the factors that influence achievement of outcomes. Instead, you might gather contextual evidence related to these factors, such as the presence or absence of supportive services such as transportation, translation or interpreter support, food programs; presence of multiple (complementary or non-complementary) services and programs within the target population or community; and children's or students' home life and family stability.

You might find such evidence in published studies or reports or refereed conference presentations, or you might turn to experts in the field, key stakeholders, and Project Officers to help you think through all of the factors that might influence project implementation or the ability of the target population to engage in and respond to

project services. However, even if you consult all of these sources, there is no guarantee you'll identify all of the potential issues your project might encounter during implementation. For this reason, it's important to build evidence from evaluation into your project plan, so you can track implementation progress, gather formative feedback to understand what's happening on the ground, and make necessary adjustments. This information can then be used to modify the theory of change, as needed, to improve current project functioning or to inform development of future projects.

How do I select a project intervention?
Key questions to ask yourself when beginning to identify project models or interventions for use in your project include:

- How can I choose an intervention that will be effective?

- How can I be confident that the intervention I'm proposing will actually result in the desired outcomes (i.e., will respond to the need)?

The first step in answering these questions is to have a thorough understanding of the need, as outlined *previously*. The next step is to determine whether a project model or specific intervention exists that has been proven successful for the specific need and the targeted population. There are a growing number of searchable clearinghouses, registries, and databanks that catalog interventions that respond to specific

needs and that describe the effectiveness of services for a target population and show the rigor with which the intervention was tested for efficacy (see below for examples). These online sites will help you determine what models or interventions meet or exceed your expectations for supporting evidence (e.g., showing *"Evidence of Promise"*).

For example, the WWC produces intervention reports, practice guides, and single study reviews to serve as primary sources of information on interventions for education policymakers, practitioners, and researchers.

- **Intervention reports** summarize findings from WWC-reviewed studies related to a particular intervention published during a certain time period. Studies that meet *WWC standards* are given an intervention rating, and the size of effects and the extent of evidence regarding the intervention's effectiveness are described.

- **Practice guides** provide recommendations for educators on a particular topic, based on reviews of research and the expertise and professional judgments of an expert panel of nationally recognized researchers and educators.

- **Single study reviews** assess findings on an intervention described in one research study. These single study reviews often result from a particular study or intervention getting considerable media attention.

Selected Research-Based Clearinghouses and Databanks

➤ **What Works Clearinghouse (WWC).** Contains a section for *Children and Youth with Disabilities* and rates programs and interventions in categories such as general academic achievement or social-emotional development. The ratings range from No Discernible Effects, Mixed Effects, Potentially Positive Effects, and Positive Effects.

➤ **California Evidence-Based Clearinghouse for Child Welfare.** Contains a databank of rated programs that may be of interest to some projects, such as Infant and Toddler Mental Health Programs or Behavioral Management Programs for Adolescents in Child Welfare. Ratings range from Not Able to be Rated, to Concerning Practice, Evidence Fails to Demonstrate Effect, Promising Research Evidence, Supported by Research Evidence, and Well-Supported by Research Evidence.

➤ **Career and Technical Education (CTE) Clearinghouse: Serving Special Populations.** This clearinghouse contains a range of products including technique articles, issue briefs, research reports, and journal articles.

➤ **Promising Practices Network.** The network databank identifies programs that have some evidence proving effectiveness. Programs can be searched by outcome area, indicator, topic, evidence-level, as well as alphabetically. Listed programs are rated as Proven, Promising, or Other Reviewed Programs.

➤ **National Registry of Evidence-Based Programs and Practices.** Sponsored by the Substance Abuse and Mental Health Services Administration, this clearinghouse rates programs as either Programs with Ineffective Outcomes, Programs with Promising Outcomes, or Programs with Effective Outcomes. The registry focuses on programs of interest or use for children and students with mental health and related needs.

As you review different clearinghouses, databanks, and academic publications, you will find that there are two general types: **packaged interventions** and **models that are collections of strategies or techniques.** There may be evidence for both types; however, the level of evidence and existing use of evidence in the form of implementation guidelines and protocols will likely be different.

The first type, **packaged interventions**, typically undergo a development and testing process to ensure that the structure, activities, and protocols all contribute to a desired outcome or outcomes. These types of interventions often contain guidelines for implementation—some of which are evidence-based—that include topics such as:

• Service provider qualifications (e.g., education, experience, and training), recommended staff supports such as supervision and technical assistance, and other desired skills that will enhance service delivery.

• Population characteristics, or details about the population for whom the program has been shown to work.

• Implementation criteria, such as required mode and dosage of services, intensity of services, minimum and maximum case load per staff, required and recommended materials, sequencing of services, and so forth.

• Recommended assessment techniques or guidance and tools for conducting evaluations of how well the project is being implemented (i.e., a formative

evaluation) and the extent to which the project contributed to desired outcomes (i.e., a summative evaluation).

These details are very helpful when planning an evaluation for an intervention. The published research (e.g., academic publications and clearinghouse summaries) also can be good sources to discover implementation details. In addition, a team can contact the intervention publisher or design team to obtain this information.

For the second type, **models that are collections of strategies or techniques**, there is generally less research evidence. However, initiatives such as the Department's *Investing in Innovation (i3) program* have been gathering data on a variety of such education programs and projects since 2010, with evaluation results starting to become publicly available.

If no program or intervention already exists to attend to the need you are addressing, you'll need to look for research into similar types of projects and gather contextual and experiential evidence from key stakeholders and experts in the field to create a plausible program model that can work to address needs.

If you do need to conduct your own search of published research to determine the level of evidence supporting a particular program, strategy, or intervention, it's important to note study details such as design, target population, implementation criteria, and study findings or outcomes. See the box below for various factors you should consider when reviewing the research evidence.

Factors to Consider When Assessing the Quality of Research Evidence

➤ Did the study demonstrate that the intervention or strategy achieved the intended or desired **effect**? If so, to what degree and how do you know (e.g., What was the effect size)?

➤ Was the study designed with acceptable **internal validity**— can the outcomes that are reported be reliably connected to the intervention or strategies being studied? Are there other plausible explanations for the observed outcomes?

➤ Was a rigorous **research design** used to observe or measure outcomes? Was outcomes testing based on a reasonable sample size? Was there a treatment and a comparison or control group?

➤ Were the same or similar outcomes identified in multiple, **independent replications** of a study? Is there at least one study by a study team other than the developer or publisher?

➤ Is there published **implementation guidance** for implementing the intervention or strategies? Can sufficient details about implementation be gleaned from published studies? Are fidelity measures available?

➤ Is there any information available about **factors or influences** that may affect project implementation or a client's ability to fully engage in and respond to the strategies or intervention?

➤ Are study findings and outcomes robust to a wide variety of contexts and populations **(external validity)**?

➤ Did the study examine the intervention or the strategies under conditions similar to those that would be found in "real life" **(ecological validity)**?

Source: Puddy & Wilkins (2011).

It's possible that not all of the dimensions of research quality will apply to your project. For example, if you have a Personnel Development (PDP) project, there may be no prior studies of the particular approach you are using to train your teachers and related-services professionals, let alone any independent replications of those studies. Or, if you have a Technical Assistance and Dissemination (TA&D) project, since there may not be a specific intervention that you are using to provide targeted or intensive TA to your stakeholders, it is unlikely that implementation guidance will exist for your particular TA approach.[5]

What's important here is that, whenever possible, you should consider these dimensions when judging the quality and extent of the evidence that exists related to a particular intervention or strategy. If the answer to most of the questions about the quality of research evidence that apply to your project is "Yes," there's probably good evidence in support of the intervention or strategy.

However, even when the research evidence is good, you should always carefully consider the alignment of an intervention or strategy with the target population (such as whether the intervention would be considered by key stakeholders to have social validity[6]) and the identified needs for your particular project.

It's important to note that a collection of strategies and activities may lack the cohesive and integrated structure of a packaged intervention, but this approach may provide greater flexibility to respond to the specific needs of your target population or context than a packaged intervention. In some cases, you might want to use a packaged intervention along with implementation of separate strategies and activities designed to address specific needs. The table below outlines some pros and cons of using packaged interventions compared to collections of strategies or techniques.

Pros and Cons of Using Packaged Interventions vs. Collections of Strategies or Techniques

Packaged Intervention	Collection of Strategies or Techniques
Pros	
➤ The development process typically establishes evidence for the intervention's success which may make it easier to implement the intervention with fidelity.	➤ A project can create a unique collection of strategies and techniques that is customized to meet the needs of the target population.
➤ The development team may publish helpful guidance documents such as implementation guidelines and evaluation instruments, which may make it easier to implement the intervention with fidelity.	➤ Unlike packaged interventions, these strategies and techniques are not likely to be proprietary, thus avoiding some costs.
Cons	
➤ The user should adhere to the established service model—this limits the project team's ability to modify or customize the program for a specific population.	➤ There may not be as much evidence supporting the use of different strategies and techniques for the purposes and population of interest.
➤ It can be costly to purchase and become trained on a packaged intervention or to purchase and use intervention instruments and documents.	➤ There may be a lack of guidance and protocols for implementing the project. These would need to be developed by the grantee, which can be a costly and time-intensive process.

---

[5] The U.S. Department of Education's Office of Special Education Programs (OSEP) has created a Vision for the Technical Assistance & Dissemination (TA&D) Network that can be used to guide provision of targeted and intensive technical assistance. OSEP expects TA&D projects to implement , whenever possible, "a set of evidence-based practices within a defined implementation strategy that will improve child or system outcomes. As used here, "evidence-based" refers to findings from research that indicate either "strong" or "possible" evidence of effectiveness of a practice. Different research techniques are capable of yielding either "strong" (e.g., well-designed randomized control trials, comparison studies using well-matched groups) or "possible" evidence of effectiveness (e.g., pre-post studies, comparison-group studies in which the intervention and comparison groups are not well-matched; and "meta-analyses" that combine the results of individual studies which do not themselves meet the threshold for "possible" evidence). Source: *https://www.osepideasthatwork.org/ sites/default/files/documents/ConceptFrmwrkLModel+Defs2012.pdf*

[6] "Social validity" refers to the social importance and acceptability of a particular project's or intervention's goals, procedures, and outcomes. Shaver, D., Wagner, M., Nagle, K., & Ryan, T. (2015). *Improving implementation of programs and practices for children with disabilities. Lessons learned from the Model Demonstration Coordination Center*. Menlo Park, CA: SRI International.

Differences between Logic Models and Theories of Change

Logic Model	Theory of Change
➤ Tends to be linear	➤ Does not have to be linear; may contain feedback loops and iterative processes
➤ Incorporates features specific to the project's investments	➤ May account for confounding factors including context, mediators, and moderators
➤ Tries to make a complex process simple	➤ Tries to present the complexity of a process or series of processes

How do I create a logic model to enact the theory of change?

A logic model is a visual representation of and organizational structure for a theory of change (also sometimes called a theory of action).[7] Building upon the theory of change, a logic model provides specific detail about the mechanisms by which the project will achieve the desired outcomes. The Appendix includes the Department's Office of Special Education Programs, *OSEP, logic model outline*, as well as a *simplified example of a logic model for a hypothetical Parent Resource and Technical Assistance (TA) Center*.

While there is no one "correct" way to create a logic model, certain elements are commonly found in logic models:

- **Inputs** include the resources that are available to the project. This includes external funding, internal resources, and intangibles such as experience and the state of the knowledge in the field.

- **Activities** are the specific actions funded by the project or supported by other resources under the umbrella of the project. Strategies are broad approaches to addressing the goals and generally include multiple activities.

- **Outputs** are the direct results of the project activities, including project products and programs. Most outputs will be quantifiable, including tallies of the number of products and programs or counts of the customer contacts with those products and programs.

- **Short-term, intermediate, and long-term outcomes.** Results, or project outcomes, are the changes that occur in the population served, as a result of your project.

We discussed previously how you might identify the available resources and *select a particular intervention* for your project. Here, we'll discuss how you might use evidence to choose strategies or activities (discussed next) and *identify realistic outputs and outcomes*.

Help! I can't find evidence in support of the strategy or intervention I'd like to use. What do I do?

Funders who are explicit about funding research-based programs will likely not want to fund a project that features an intervention or strategies that don't have research support. In these instances, you might want to request a conversation with the funder to discuss acceptable options. Or, if possible, you might seek additional funders interested in innovative practices. Programs such as the Department's *Investing in Innovation (i3)* Program recognize the value of assessing the effectiveness of educational projects that do not yet have a strong research base, so they have created tiered evidence programs that offer funds to develop innovative projects based on the scope and level of supporting evidence.

How do I select specific strategies or activities for my project?

It's more and more common for funders and sponsoring agencies to require strategies and activities that are research-based. This means that there is high-quality data from rigorous research studies (such as those that meet *WWC standards*) that supports the use of a strategy or activity to respond to a particular need and result in a desired outcome.

If you're using a packaged intervention that has a strong research base, it's likely that implementation strategies and activities have already been identified by the program developer. If you're using a combination of strategies or techniques as the project intervention, however, you'll need to look for evidence that the approach you intend to use can reasonably be expected to produce the outputs and result in achievement of the project outcomes.

Whichever approach you take, we encourage you to find supporting evidence for the strategies or activities you'd like to use. This may involve conducting internet and other searches—such as the academic base of published

---

[7]   Frechtling, J.A. (2007). *Logic modeling methods in program evaluation.* San Francisco: Jossey-Bass.

studies (such as *Education Resources Information Center (ERIC)*) or through digital libraries such as *JSTOR* to identify promising practices, strategies, and activities. It's always advisable to explore the academic base for different strategies and activities to gain an appreciation for the different ways a specific element may have been studied. In particular, it's helpful to understand how one strategy may have been deployed across different populations, communities, and contexts—and to determine if the same level of effectiveness can be established in each case. Similarly, you might talk with other project teams who are working in your field or with local stakeholders to gather contextual evidence and experiential evidence to tell you whether the chosen strategies and activities are likely to work in your context and lead to the desired results. In doing this, it's important to carefully think through how you believe the specific strategies and activities, in which amounts and in which contexts, can be expected to produce your desired results.

How do I identify realistic outputs and outcomes?

Another way you might use evidence at the Planning Phase is to identify realistic outputs and outcomes for your project. This is an important part of planning, because you need to think through what your project can truly be expected to accomplish during the project period. There are several ways you might do this. First, you can refer to your funder or sponsoring agencies' reporting or evaluation requirements to determine what product, service, and client count data must be reported (e.g., such as program performance measures) and which outcomes would be viewed as signs of success or progress. Second, you can turn to the research evidence and intervention materials (if using a packaged intervention) for ideas on the types of products, services, and client counts that have been tracked through similar projects and to find the outcomes that have been studied and use these same outputs and outcomes (and measurement techniques) in your project. Finally, you can meet with project stakeholders and advisors to determine which outputs and outcomes are of highest priority and interest (experiential evidence), as well as to learn whether within a given context it might truly be possible to produce certain outputs and outcomes (contextual evidence). For example, it may not be realistic to expect that a TA&D project could implement a series of 10 full-day training workshops for teachers during the school year, given the fact that it is often hard for schools to secure substitute teachers. Similarly, a PDP project focused on preparing teachers realistically could not be expected to demonstrate how it could achieve a long-term outcome of improving college completion rates among deaf and hard of hearing students taught by these teachers.

How can I incorporate evaluation evidence into the project?

Evaluation should be an integral part of your project; therefore, a comprehensive evaluation plan should be developed at the same time as—or as part of—the overall project plan (see below for common elements in an evaluation plan). Generally, the purposes of an evaluation are to demonstrate how well the project components have been implemented and to analyze the extent to which the

Common Elements in an Evaluation Plan

➤ **Introduction.** A review of the project and its theory of change—often represented as a logic model—as well as contextual factors to be considered during the evaluation.

➤ **Evaluation questions.** Commonly, an evaluation seeks to answer questions about a project's overall effectiveness and efficiency, as well as respond to specific questions defined by a funder. An evaluation may contain both formative and summative evaluation questions. Formative questions focus on the extent and quality of project implementation while summative questions focus on the extent to which a project achieved its goals (i.e., outcomes).

➤ **Methodology.** Specific details regarding the evaluation design, data collection, data entry and management, data analysis, and reporting.

➤ **Timeframes and responsibilities.** A timeline for the overall evaluation project as well as timelines for specific evaluation events such as data collection, analysis and reporting. This section also can identify the project staff that will complete different evaluation tasks and include the roles and responsibilities of the third-party evaluator.

➤ **Deliverables.** The reports and other products to be generated from the evaluation.

➤ **Budget.** The costs of conducting the different evaluation activities, including staff hours, travel costs, materials, etc. Third-party evaluators often will include a budget in their proposed evaluation plan. Even if you don't work with a third-party evaluator, you should consider the resources you might have available for evaluation.

project's objectives and outcomes have been achieved. The results of such evaluations provide project implementers with evidence to make decisions about project improvements, expansion, and sustainability; assess efficiency and guide cost-containment strategies; and facilitate replication in other settings. More importantly, evaluation results can provide information on a project's impact—information that can be used by the funder and by other key stakeholders to make an assessment of the scope and value of project achievements. OSEP grantees are required to report on their project's accomplishments using tools such as the Annual Performance Report.

If a project element is important enough to be in the logic model, it typically is important enough to be included in the project evaluation. In brief, logic models assist in the following evaluation planning tasks:

- Establishing the linear, or chronological, sequence of evaluation events. By grouping logic model elements in a particular linear sequence, which can then be aligned with specific evaluation components—such as progress monitoring, formative feedback and summative evaluation—the project is establishing which changes are expected to happen first, second, and so forth.

- Identifying important project measurement targets— these are the project components and products that will be captured in the evaluation, such as:

  - Formative evaluation, which includes the activities used to monitor and assess the numbers and types of services provided, products delivered, and clients served, as well as the accessibility, quality, and relevance of the program for client needs; and

- Summative evaluation, which includes the activities used to assess whether and to what degree change occurred in the target population as a result of program activities. Summative evaluation also captures the conditions under which change was made possible.

This will help you to map out how you will evaluate the various parts of your project.

When available, the theory of change also can assist the evaluation by establishing whether or not change is expected to happen in a linear way or if, in theory, change is expected through an iterative or circular process. Further, the theory of change can alert the evaluator as to conditions that influence whether and how well a program can be implemented (i.e., mediators) and whether or not different populations of clients will be experiencing change in the same way or at all (i.e., moderators).

The box below presents some important factors to consider when planning an evaluation. For more information on how you can design and conduct a high-quality evaluation for your project, we suggest you see the *Evaluating Special Education Projects: Resource Toolkit*, which can be found on the OSEP IDEAs That Work website. It discusses the steps to planning and conducting a project evaluation, outlines some methodological considerations, and offers resources where you might find more information on the topics.

Now, let's turn to how you can discover and use evidence during the Implementation Phase. We will continue to discuss the importance of evaluation evidence throughout.

Considerations for Evaluation Planning

A key part of evaluation design is the **rigor or intensity** of the evaluation. Consider the following when deciding "how much" evaluation your project should (or can) invest in:

➤ **Evaluations require an investment from project staff,** even when the evaluation is outsourced to an external team. Consider which of your project staff will have the background and availability to work on the evaluation, for the level of evaluation that is required. CIPP developed *Guidelines for Working with Third-Party Evaluators* for those project staff interested in bringing on an external evaluator.

➤ **Evaluations take time.** A project evaluation begins the day the project is funded and can continue past the date project funding ends. Consider the inclusion of evaluation activities throughout the project's lifespan.

➤ **High-quality data are valuable.** High-quality data create impactful evaluation findings. That stated, high-quality data may be expensive to acquire. The use of standardized tools and measurement techniques may require the purchase of materials and specialized training.

➤ **The presence of a control group or well-matched comparison group increases the rigor of an evaluation.** The incorporation of control or comparison groups also will likely increase the cost of an evaluation. Choose the most rigorous design your project can afford, for the level of evidence you are expected to produce. For example, if you are expected to prove the unique contributions of your program to a desired outcome or set of outcomes, you may need a more rigorous design. In contrast, if the expectation is to correlate the program to observed results, a less rigorous design may be reasonable. CIPP created a toolkit of evaluation resources that includes detailed information on planning and conducting an evaluation, including information on different types of research designs that might be appropriate for OSEP projects.

Implementation Phase

This section describes what and how evidence can be used to help ensure the effective implementation and evaluation of your project plan and will cover:

- ➤ What are formative evaluation data and what they can tell project staff
- ➤ How to collect and use data to monitor project progress, including:
  - Describing progress monitoring data and how to collect them
  - Using progress monitoring data to keep the project on track
- ➤ How to collect and use formative data to inform and assess project implementation, including:

- How to know if implementation is going well, including:
- Using formative data to inform project functioning and guide improvements
  - – Determining if the project is being implemented with fidelity
  - – Determining if the project is on track to achieve expected outcomes

What are formative evaluation data and what can they tell me?

At the most basic level, projects don't achieve their desired outcomes because of breakdowns in the planning or implementation processes.[8] Collecting and using evidence about implementation can help you (1) monitor whether activities are carried out and outputs are generated in a way that is likely to result in expected outcomes; (2) understand why your project is failing or succeeding; (3) make informed decisions about project improvements that may, in turn, prevent failure or enhance success; and (4) aid in the documentation of project implementation to inform future dissemination, replication, or expansion efforts.

At this point in the project, you will need to carry out specific data collection activities to gather evidence related to your project. In general, evidence produced by formative evaluations addresses how well the project or intervention is being implemented, including the nature of the activities, products, and support provided by project staff; the structures, policies, and procedures influencing implementation and outcomes (e.g., contextual evidence); fidelity to the project model; changes that may be necessary to improve project implementation; the ways the project is being perceived by key stakeholders (e.g., experiential evidence); and progress toward achieving outcomes.

Discovery and Use of Evidence in the Implementation Phase

What Do I Need?	How Might I Use It?
➤ Evidence in the form of formative evaluation data on • Project progress • Implementation of activities • Project outputs and short-term outcomes • Fidelity	➤ Make mid-course corrections ➤ Monitor progress toward achieving outputs and outcomes ➤ Identify barriers or facilitators to implementation ➤ Assess level of fidelity

[8] Patton, M. Q. (2008). *Utilization-focused evaluation, 4th edition.* Sage: Thousand Oaks, CA.

Formative data can come from a variety of sources and be collected through many methods. Formative data are generally collected to answer questions that relate to:

- **monitoring progress** toward carrying out activities, producing outputs, and achieving the short-term outcomes identified in a project's logic model;

- **social validity**—the social importance and acceptability of the project or intervention, such as the social significance of the project or intervention goals, the social appropriateness of the intervention procedures, and the social importance of the intervention outcomes;[9] and

- **fidelity**—in this phase fidelity data relate to whether and to what degree the project is carrying out the strategies and activities and producing outputs as intended (i.e., in the expected amounts and covering the expected content).

Progress monitoring is generally part of a broader formative evaluation that is designed to tell project staff and other key stakeholders to what degree and how well a project is being implemented. It has a narrower focus than formative evaluation however, so we'll discuss it separately.

How do I collect and use formative data to monitor progress?

Progress monitoring is "the systematic and continual documentation of key aspects of program performance that assesses whether the program is operating as intended or according to some appropriate standard."[10] Progress monitoring helps project staff and evaluators to know whether activities are proceeding as planned and can signal the need for increased effort or changes to project implementation. During the early stages of project implementation, progress monitoring may serve as the primary gauge of whether the project is moving towards achieving its objectives. Monitoring is essentially oversight of the project's implementation phase and its purpose is to track project functioning. This includes the initial implementation activities such as preparation for data collection or provision of services, as well as tracking implementation throughout the project period. Some questions you can answer through this process are:

- Is the project achieving milestones and benchmarks in a timely manner?

- Is the project in compliance with the federal priorities?

- Is project staffing sufficient in numbers and competencies?

- Are resources adequate to support project activities?

- How many persons have participated in activities or received services?

What are progress monitoring data and how do I collect them?

Data for progress monitoring are **primarily quantitative** and include metrics and benchmarks related to project implementation, such as the number and types of participants that the project should reach each year or the number and types of services delivered and received by the target audience. Progress monitoring data might include information about the delivery, content, and attendance of participants at project training or coaching sessions; and associated expenses for training or coaching sessions. Additionally, progress monitoring data can be used to assess efficiency and guide cost containment strategies, if needed. Depending on the nature of the metrics and benchmarks, progress monitoring may be helpful for the completion of Annual Performance Reports, or other required reports.

Possible Sources of Formative Data

- ➤ **Documents or products** such as schedules and agendas, meeting notes, attendance sheets, and products produced as outputs;

- ➤ **Surveys or interviews** of stakeholders that include questions about the quality, relevance, usefulness or social validity of project activities; changes in learning or action as a result of project activities; or fidelity to the project model (For more information on how to conduct surveys and interviews, see the CIPP webinar series on *Customer Surveys* and *Qualitative Interviews*);

- ➤ **Observations** of stakeholders in their practice settings (e.g., either before and after training or only after training); and

- ➤ **Assessments**—either pre-post assessments of change in individual stakeholders or groups, or, preferably, assessments that compare stakeholders who participate in project activities with those who do not.

9 Shaver, D., Wagner, M., Nagle, K., & Ryan, T. (2015). *Improving implementation of programs and practices for children with disabilities: Lessons learned from the Model Demonstration Coordination Center*. Menlo Park, CA: SRI International. Retrieved from *http://mdcc.sri.com/documents/MDCC_Final_Report_SEPT2015.pdf*

10 Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach, 7th edition*. Sage: Thousand Oaks, CA.

We recommend defining benchmarks and developing questions to guide progress monitoring as early as possible in the project planning process.

Progress monitoring data can be collected through sources such as schedules and agendas, meeting notes, attendance sheets, expert reviews, and stakeholder observations, and they are generally collected via a tracking database or management information system (MIS). The box below includes some factors to consider when developing a data tracking system.

How can I use progress monitoring data to keep my project on track?

Ultimately, what progress monitoring data you collect should be guided by your project's logic model and project plan—including your evaluation plan. As discussed in the section on *Planning*, a *logic model* shows the connections between the major project components (i.e., inputs, activities, outputs, and outcomes).

The project plan should further operationalize these components by providing details such as who, what, when, where, or how much. Taken together, you can use this information to pinpoint the critical resources (e.g., staffing, funding); conditions (e.g., timing, location, numbers to be served); and actions (e.g., recruitment, training) that are necessary for your project to accomplish its activities.[11] You can then develop indicators and targets for each critical project component to assess whether the project is functioning as intended. These indicators are different from program and project performance measures, but they can provide important information to help you know whether you are on track to achieve the performance targets. The figure on the next page shows an example of how progress monitoring data can be used to alert project staff and evaluators whether project activities are on track. In the table, color coding is used to illustrate the status of project activities and whether they are ahead of schedule, on schedule, or behind schedule.

Factors to Consider When Creating a Data Tracking System

Data tracking systems can be basic spreadsheets operated and maintained by a single person or sophisticated web-based databases such as management information systems (MIS). An MIS is a computerized system that compiles information from inside and outside an organization. These data might relate to resources, staffing, clients, or project components. The MIS is used to process, integrate, and store these data into a centralized database where it is constantly updated and made available to all authorized users in a form suited to their purpose. MIS are often accessible to multiple users spread out across many sites. Typically, project staff work with an information technology specialist to develop an MIS.

The type of system you choose will depend on the size and complexity of the project and the available resources. In general, when making decisions about which type of tracking system you need, you should ask yourself five questions, the answers to which will depend in part on the types of data that need to be included in the tracking system:

➤ What kind of database do I want to use? (e.g., a simple database such as Microsoft Excel, or a relational database such as Microsoft Access);

➤ How do I want to input the data? (e.g., manual entry by a project team member, or a web form linked to the database);

➤ How will I verify the data? (e.g., processes for data verification and cleaning);

➤ How do I want to update the data? (e.g., point-in-time through manual entry, or real-time through web forms); and

➤ How do I want to analyze the data? (e.g., from within the database using queries and reports to generate descriptive analyses, or by exporting the data to a statistical analysis program such as SPSS/SAS/STATA to conduct descriptive and inferential analyses).

---

[11] Rossi, Lipsey, & Freeman, 2004.

	Last Updated 5/17/2016				Actual Ahead of Plan
		Not Started			Actual Behind Plan
		On Schedule			Actual = Plan
		Complete			Not Complete
		At Risk			

Activity	Notes	Status	Planned Begin	Planned End	Actual Begin	Actual End
Data Collection Prep						
Instrument development						
Develop draft protocols		Complete	3/7	3/17	3/7	3/16
Obtain external feedback on draft protocols		Complete	3/18	3/25	3/17	3/22
Incorporate feedback into revised protocols		Complete	3/25	3/30	3/25	3/28
Pilot protocols	Pilot took longer than expected	Complete	3/31	4/13	3/31	5/3
Finalize protocols		At Risk	4/13	4/18	5/3	
Institutional Review Board approval						
Submit original IRB application (with consent forms)		Complete	2/1	2/26	2/1	2/26
Obtain original IRB approval	Still waiting for approval of consent forms	At Risk	2/29	3/18		
Submit amendment to IRB with all final instruments		Complete	4/1	4/1	4/25	4/25
IRBs approve amendment		Complete	4/4	4/8	5/3	5/11
Site Selection and Recruitment						
Obtain list of potential sites		On Schedule	5/1	5/21		
Select sites		On Schedule	5/22	5/25		
Send introductory email to sites		Not Started	5/28	6/1		
Obtain principal/site leader approval forms		Not Started	6/4	6/8		

Tools like the one in the figure above can be very helpful and can be created using a spreadsheet program such as Microsoft Excel. More sophisticated management information systems can also be used (see the box on the previous page for factors to consider when creating a data tracking system), and those can even be programmed to send notices to project staff when key timelines are approaching or have passed without an activity being completed. Similarly, many technical assistance (TA) centers use databases to track the TA delivered to stakeholders. These databases often include information such as the date of the initial TA request, the nature of the request, the timing and content of the initial response from the TA provider, any follow-up responses, and the date of completion of the TA request.

In order to be useful, you should collect and analyze data related to progress monitoring indicators on a regular basis. The table on the next page presents some sample indicators for a hypothetical Parent Resource and Technical Assistance (TA) Center and suggested actions you might take to address unmet indicators (see the Appendix for the *sample logic model* for the Center).

Sample Progress Monitoring Indicators and Actions to Address Unmet Indicators

Progress Monitoring Indicators	Action If the Indicator Is Not Being Met
1. Training workshops identified in the project plan are conducted and occur within the month specified in the project plan timeline.	➤ Meet with project staff to determine why and identify what actions can be taken to help get the workshop schedule back on track. This might require working with the local school district to ensure facilities are available during the scheduled workshop times.
2. Staff are within 10% of the monthly hours budgeted for them on the project (e.g., if a staff member is budgeted to work 120 hours a month then they should log between 108-132 hours).	➤ Meet with staff member to discuss why and determine if and how to adjust staff responsibilities. If it is not possible to adjust responsibilities, it might be necessary to reallocate hours from another staff member or task to cover the extra hours needed.
3. Families from all targeted zip codes have attended at least one training workshop (start monitoring after three months of implementation).	➤ Meet with project staff and key stakeholders to discuss why and identify how to encourage more participation. This might involve offering the workshops in different locations or at different times, or conducting additional outreach to encourage parent participation.

If you see that project monitoring indicators are not being met, a first step is to meet with project staff and possibly key stakeholders to explore the research, contextual, and experiential evidence related to the issues, identify possible solutions to problems, and then make decisions about project adjustments accordingly. Consistent with the available evidence, action should be taken to address any problems to ensure the project stays on track and has a good likelihood of achieving its expected outcomes. If it becomes clear that you won't be able to address all of the identified problems with project functioning during a certain time period, you should talk with project staff and key stakeholders to determine what changes can be made so as to increase the likelihood that you will achieve as many of your outcomes as possible. You may also consider whether it is necessary to adjust your project's theory of change if the data you've collected through your evaluation shows that previously unexpected factors might be influencing project functioning. Then, you should update your logic model accordingly.

How do I collect and use formative data to inform and assess project implementation?

As discussed earlier, the project logic model can serve as your guide as you gather data to track progress toward implementing activities and producing outputs and outcomes.

Generally, the data you'll collect during this phase relate to three aspects of project implementation:

• outputs

• short-term (and for some projects intermediate) outcomes

• fidelity

Outputs are the direct result of project activities and they play an important role in understanding project implementation, as they are integral to both progress monitoring and assessing fidelity. They are an early and concrete source of information to consult when examining project implementation. Output data can tell you which activities have occurred, how many people participated in them, what services have been provided, and what products have been developed. They also can provide the information necessary to determine "dosage" or the amount of services received by participants.

Short-term outcomes are what are expected to occur as a result of the outputs. They are the immediate benefits that participants are anticipated to experience or display as a result of project activities. These outcomes often relate to changes in participant attitudes, knowledge, or skills. Ideally, you would collect these data from participants before and after the delivery of project services and then analyze the data to determine if the expected changes occurred (and if not, you might gather social validity data to figure out why).

Fidelity data tell project staff and evaluators the degree to which an intervention or project is being delivered as intended.[12] The fidelity data you might want to collect during the Implementation Phase focus on whether the project activities and outputs are being carried out as intended, changes in learning (short-term outcome) and, in some cases, changes in actions (intermediate outcome) among key stakeholders.

How do I know if project implementation is going well?

Some questions you can ask to determine whether project implementation is going well are:

- Is the project completing planned activities and producing the expected outputs?
- Are there any implementation gaps or project support needs?
- Are there any facilitators of or barriers to implementation?
- Is the project being implemented with fidelity?
- Do all signs point to achieving desired outcomes?

The box below outlines some possible scenarios you might encounter as you assess project implementation as well as some questions you might ask to figure out why the data might be showing that your project is not proceeding as planned.

Monitoring Achievement of Outputs and Short-Term Outcomes

Analyses of formative data will likely indicate one of the following:

➤ **Scenario 1**
Your project is proceeding as planned and achieving (or exceeding) projected levels of implementation and progress toward expected outcomes,

➤ **Scenario 2**
Your project is proceeding as planned and achieving (or exceeding) projected levels of implementation, but no progress has been made toward expected outcomes, or

➤ **Scenario 3**
Some combination of Scenarios 1 and 2—parts of your project are being implemented as planned and you are achieving mixed success with short-term outcomes.

What do these scenarios tell us about the project?

Scenario 1 is optimal and suggests that no major revisions to project implementation or the logic model or theory of change are needed at this time. **Scenario 3** is probably the most typical. There are issues with implementation and this is affecting the extent to which short-term outcomes are being achieved. If there is no relationship between the lack of progress toward outcomes and the problematic areas of implementation, then **Scenario 2** may apply (e.g., implementation is strong but the project isn't producing expected outcomes).

Scenario 2 is the most problematic. If your data show that project implementation seems to be proceeding as planned, but it is not achieving the expected short-term outcomes, what do you do? Determining the source or sources of the problem may not be easy. Some questions you might ask include:

➤ Were the data collection instruments developed by the project?

  ✦ If so, was any field testing conducted to help determine if the instruments are sensitive to the types of changes that would be expected and if they reliably measure what was intended?

  ✦ If the instruments were not developed by the project, do you know anything about the reliability and validity of the instruments?

➤ Did you collect fidelity data as well as progress monitoring data?

  ✦ If yes, has the project been implemented with fidelity to the model across all participating sites?

➤ Are there factors that affect the ability of participants to fully receive or respond to project activities or services (e.g., such as if the intervention has low social validity)?

  ✦ If yes, are these factors included in your logic model or theory of change, or might it be necessary to make adjustments to the project logic?

---

[12] Carroll, C., Patterson, M., Wood, S, Booth, A, Rick, J, & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science. 2*(40).

How do I use formative data to inform project functioning and guide improvements?

Formative data are frequently used to describe how projects are unfolding, including the degree to which they are meeting the needs of various stakeholders, following the planned timeline, and operating as envisioned. These data provide information on whether there are gaps in implementation or a lack of supports necessary for the project to achieve its short-term or intermediate outcomes.

Another way to use data collected at this phase is to identify barriers to and facilitators of project success. These may be persons, policies, places, plans, etc., and are often unexpected or unintended. For example, one possible barrier to project success is low social validity.

The example below shows how formative data might be used to inform project functioning and guide improvements to implementation.

Example: Using Data to Inform Project Functioning and Guide Improvements

As part of its ongoing TA, our hypothetical Parent Resource and TA Center wants to improve the early identification of autism and increase support for children with autism and their families. One component of the project involves training preschool lead teachers to identify the early signs of autism and to provide support to families of children with autism.

Activities
Deliver a series of six teacher training workshops during the fall semester.

Data Sources
Training schedule and agendas, training sign-in sheets, observations of trainings, interviews with project staff, post-training teacher survey, online teacher survey (surveys have > 80% response rates), coaching logs, interviews with teachers getting coaching.

Findings
➤ **Implementation gap:** Only four trainings actually delivered.

➤ **Implementation gap:** Few teachers actually attended two of the four trainings.

➤ **Project support need:** Teachers need more support to understand how to identify the early signs of autism.

➤ **Implementation barrier:** Problems with scheduling.

- ◆ Project staff were unable to schedule the remaining trainings because the experts weren't available on the days the district leaders requested.

- ◆ Online survey results indicate that many teachers were unable to take time off or couldn't find substitutes for their classes on the scheduled training dates.

➤ **Implementation barrier:** Limited opportunity for practical application of learning during the trainings. Online survey results indicate teachers who do attend the trainings feel that more time for hands-on practice of the skills they are learning is needed.

➤ **Implementation facilitator:** Higher rates of coaching are associated with better understanding of the signs of autism among preschool lead teachers (one of the project's expected short-term outcomes).

➤ **Implementation facilitator:** Prior project experience and analysis of completed homework activities (which include reviewing a case study of a child's behavior and determining whether the child demonstrates the early signs of autism) illustrates that teachers are better able to apply their skills of identification of and support for children with autism (one of the project's intermediate outcomes) after completing all six training modules.

Project Actions
➤ Create six online training modules that teachers can view at a time that is convenient.

➤ Offer additional supports to teachers:

- ◆ Incorporate homework to be completed prior to and after each module to give teachers more practical application of the content.

- ◆ Offer teachers who complete the training series on-going opportunities for interaction with a coach so that teachers can apply what they are learning.

- ◆ Increase the number of coaches on the project team.

- ◆ Incorporate additional opportunities for applied homework in the training modules.

How do I know if the project is being implemented with fidelity?

Fidelity data are essential to help you understand what's happening with project implementation. To be able to link project activities to outcomes, for example, it's not enough to know whether an activity has occurred. It's also important to know

- whether the activity was carried out in the way that was intended (e.g., Did the activity get conducted in the correct timeframe and cover the expected content?);

- whether the right people attended in the right amounts (e.g., Did a high percentage of the target population attend and how often did they attend?); and

- whether the activity resulted in the expected outcomes (e.g., Were there changes in participants' knowledge, attitudes, or skills as a result of participating in the activity?).

Why Assess Fidelity?

Without knowing whether a project has actually been implemented according to plan, it's impossible to know whether the project has been responsible for producing the observed results. Measuring fidelity of the various project components can help project staff to identify "key ingredients" to project success, determine areas of improvement, and inform why desired changes may or may not be occurring.

Fidelity data are formative to the extent that they are used by project staff or program developers to make changes to a project or intervention during the course of the project. When used in a summative sense (as part of assessing results), fidelity data can offer insights into why a project might not have achieved the expected outcomes (e.g., if there was low fidelity) and signal the feasibility of implementing such a project again, in different a context. It might be good to also gather data on social validity so that you can have an idea whether problems in that area might be affecting fidelity.

Steps to Create a Fidelity Measurement System

1. **Identify the "key components" of the project, or those features that are critical for the project to achieve positive results.** These key components should be clearly illustrated in the project logic model.

2. **Operationally define each key component (e.g., professional development) included in the logic model.** An operational definition is composed of indicators, which are specific aspects of the intervention that can be measured quantitatively. The number of indicators should reflect the complexity of the project. It's best to have more than one indicator of a specific component, preferably from different sources. Additionally, it's best if the indicators are able to differentiate among different levels of fidelity. Be sure to identify indicators that can actually be measured.

3. **Select data sources and measures.** To do this, consider the research questions guiding the study, find the best data sources for each indicator, and choose whether to use existing instruments or develop new ones. Use multiple sources of data for each indicator (e.g., surveys and observations) when possible, and consider reliability/validity of data sources and measures.

4. **Establish fidelity thresholds.** Fidelity thresholds tell you the least amount of each indicator that needs to be present for fidelity to be considered "adequate." Represented as numeric scales, thresholds quantify the extent to which an indicator was enacted with fidelity. This scale can be dichotomous (0 or 1) or it can range from 0 to 3, or even 0 to 5.

5. **Calculate fidelity scores.** Fidelity scores are calculated based on the fidelity thresholds and tell you the least amount of each component that needs to be present for fidelity to be considered "adequate." You may want to calculate a fidelity score for each key component separately, or create one score for fidelity across all project components. See the Appendix for an example of a *fidelity matrix* for our hypothetical Parent Resource and TA Center.

Some packaged programs or interventions have well-defined fidelity criteria and the program developers can provide guidance on collection of fidelity data. For example, the Positive Behavioral Interventions and Supports (PBIS) framework identifies the core structural and process features at each of the three Tiers contained in its prevention model. Validated tools to measure fidelity, such as the School-wide Positive Behavioral Interventions and Support (SWPBIS) Tiered Fidelity Inventory (TFI),[13] and the PBIS Blueprints have been developed and are available on the *PBIS Technical Assistance Center website*.

Other packaged programs have less concrete (or no) fidelity criteria and only provide relatively general guidance of what an intervention should look like (even if they have very detailed implementation guidelines such as Teachers Guides). For example, a popular writing intervention includes instructions in the Teachers' Guide to adapt the lessons to provide differentiated instruction for learners of different abilities, yet gives little indication how to make adaptations and still remain faithful to the principles of the intervention (e.g., the structure and sequence of the language and the activities).

In many cases, you'll be implementing your own intervention or combination of interventions, as might be the case for a TA&D project. In these cases, you'll need to develop fidelity criteria for your specific project. The box on the previous page briefly outlines steps you can follow to create a system to measure fidelity. Once you've created your fidelity measurement system, you can collect data on each of the indicators and track how well the project is achieving them, much as you would for other evaluation activities.

How can I determine if the project is on track to achieve the expected outcomes?

As part of your formative evaluation, you'll be going through an iterative process of collecting and analyzing data to monitor progress toward completing the planned activities and producing scheduled outputs, providing feedback to improve project functioning, and tracking progress toward outcomes achievement. If you made careful use of existing evidence when you created your theory of change and corresponding logic model, you've already made a good argument that, if carried out with fidelity to the plan, your project will likely result in the eventual achievement of the long-term outcomes (e.g., improved outcomes for students with disabilities). For example, if the data produced by your project evaluation indicate that the project has been implemented with fidelity, that both the short-term and intermediate outcomes have been achieved, and that there appear to be no major barriers to project implementation going forward, it can be reasonably concluded that the project is likely to achieve its long-term outcomes.

As you transition to the Results Phase, the focus now turns more toward evaluating the degree to which you have achieved your expected outcomes and making decisions related to future project activities.

---

[13] Algozzine, B., Barrett, S., Eber, L., George, H., Horner, R., Lewis, T., Putnam, B., Swain-Bradway, J., McIntosh, K., & Sugai, G., (2014). *School-wide PBIS Tiered Fidelity Inventory*. OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. *www.pbis.org*

Results Phase
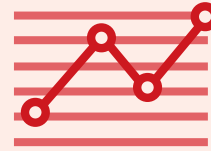
In this section, we'll talk through how evidence can help you to demonstrate your project's results, including:

➤ What are summative data and why they are needed

➤ How to collect and use summative data to show project results, including:

  • Considering factors that might influence a project's ability to demonstrate "effectiveness"

  • Determining whether the project is (or will be) successful, including:

– Knowing what to do if extensive summative data cannot be collected

– Making projections about whether the project will ultimately achieve its long-term outcomes

  • Using evidence to inform decisions related to project continuation, replication, and dissemination

What are summative data and why do I need them?

Data are considered "summative" when they are used to assess the degree to which a program has been effective in achieving its intended outcomes or to identify unintended outcomes, whether positive or negative. Additionally, summative data are used to make judgments about project quality, feasibility, or cost-effectiveness, or to make decisions about ongoing funding, replication, or scale-up. An important role for a summative evaluation is to determine the unique contribution of the project to the desired change. As such, summative questions are best informed when there are comparison data (e.g., comparison groups) that can give you an idea of the counterfactual—that is, what would have happened if the project hadn't been implemented.

Examples of summative questions include the following:

• What outcomes (expected and unexpected) have occurred?

• What expected outcomes have not occurred?

• To what degree have outcomes occurred?

• Where is change the greatest?

• What is the unique contribution of the program to the observed change?

• What is the cost/benefit of these outcomes?

• To what extent do the same outcomes occur in treatment and control groups, comparison groups, or the same group over time?

Discovery and Use of Evidence in the Results Phase

What Do I Need?	How Might I Use It?

➤ Evidence in the form of summative evaluation data related to:
  • short-term outcomes
  • intermediate outcomes
  • long-term outcomes
	➤ Assess achievement of short-term and intermediate outcomes
	➤ Determine progress toward (or likelihood of) achieving long-term outcomes
	➤ Link short-term and intermediate outcomes to projected long-term outcomes
	➤ Demonstrate achievement of meaningful project objectives
	➤ Support project continuation, replication, or dissemination

How do I collect and use summative data to show project results?

Throughout the project *Implementation Phase*, you've been collecting and analyzing data to determine whether you're on track to achieve your short-term and, possibly, intermediate outcomes. As you do so, again based on your strong *logic model*, you're also determining your progress toward (or likelihood of) achieving your expected long-term outcomes—the ultimate measures of the project's success. As we mentioned in the section on *Planning*, you should include in your project plan data sources and data collection methods that will allow you to demonstrate that the project has achieved (or is likely to achieve) its outcomes. For example, some key data you might collect in the Results Phase include:

- Pre-post assessments of treatment group knowledge, skills, actions, etc.;
- Data on possible alternate explanations (or confounds) for the observed results;

- Data related to "business-as-usual" (or the counterfactual);
- Assessments of control/comparison group knowledge, skills, actions, etc.; and
- Fidelity data related to the degree to which the project has resulted in the expected outcomes.

Like formative data, summative evaluation data may be collected via assessments, surveys, etc. (see the box below), but the difference is that these data are used to assess whether and to what degree the project has resulted in the intended outcomes.

It's a good idea to identify the possible answers to the summative questions you might test through your evaluation. This will help ensure the data collected can be used to answer those questions. The example on the next page shows how data might be used to answer evaluation questions for our hypothetical Parent Resource and TA Center.

Possible Sources of Summative Evaluation Data

➤ **Project products** intended to lead to changes in learning, action, or conditions among the target population;

➤ **Surveys or interviews** of stakeholders that include questions about social validity, or changes in learning (awareness, knowledge, attitudes, skills, opinions, aspirations, or motivations) or action (behaviors, practices, decision making, policies, or social action or organizational change) **(For more information on how to conduct surveys and interviews, see the CIPP webinar series on** *Customer Surveys* **and** *Qualitative Interviews***)**;

➤ **Observations** of stakeholders in their practice settings (e.g., pre- and post-training or, especially if you have a comparison group whose performance you can compare to that of project participants, post-training only);

➤ **Assessments**—either pre-post assessments of change in individual stakeholders or groups, or, preferably, assessments that compare stakeholders who participate in project activities with those who do not; and

➤ **Administrative data**—such as state tests or teacher certification records.

It's important to note that to answer any given evaluation question you might use more than one analytic or statistical technique. The choice of technique depends upon the nature of the (a) evaluation design and (b) summative data that are collected. Generally, experimental and quasi-experimental designs, or designs with both treatment and control or comparison groups, are considered the strongest designs for making determinations of impact and calculating magnitude of effects.

As such, whenever possible, we recommend you collect data from those who received project services or supports and from those who did not. When collected through an evaluation that features a rigorous research design such as a high-quality quasi-experimental design that features matched treatment and comparison groups, such data help you to determine the true contribution of your project toward achieving the observed outcomes. If you're not able to collect data from a comparison group, your next best option is to collect data from project participants before and after project activities.

This will allow you to see whether participation in project activities is associated with changes in outcomes.[14] Another way to demonstrate results without the inclusion of a comparison group is to use a single-case design, especially one that incorporates randomization.[15] Unfortunately, it's beyond the scope of this tool to discuss research design; however, many very good and comprehensive resources do exist. For example, we suggest you consult the Evaluating Special Education Projects: Resource Toolkit, which can be found on the OSEP *IDEAs That Work* website; Shadish, Cook, & Campbell (2002)[16]; and the *WWC standards*.

Example: Collecting and Using Data to Answer Evaluation Questions

Evaluation Question	Possible Answers to Question	Data Sources	Sample Analytic Approaches
Do preschool lead teachers show increased ability to recognize the early signs of autism in young children (short-term outcome)?	➤ **Answer 1:** There is no change in teachers' ability to recognize the early signs of autism ➤ **Answer 2:** Teachers show increased ability to recognize the early signs of autism ➤ **Answer 3:** Teachers show decreased ability to recognize the early signs of autism	Teacher surveys and knowledge assessments	➤ Compare results of teacher surveys and knowledge assessments for treatment and control/comparison groups ➤ Compare pre-post change for treatment teachers
Are preschool lead teachers able to apply their skills of identification of and support for children in their practice (intermediate outcome)?	➤ **Answer 1:** There is no change in teachers' ability to apply what they learn about the early signs of autism in their practice ➤ **Answer 2:** Teachers show increased ability to apply what they learn about the early signs of autism in their practice ➤ **Answer 3:** Teachers show decreased ability to apply what they learn about the early signs of autism in their practice	➤ Pre-post teacher skill assessment ➤ Review of teacher support plans ➤ Classroom observation	➤ Compare results of teacher skill assessments for treatment and control/comparison groups ➤ Compare pre-post change for treatment teachers

---

What factors might influence my project's ability to demonstrate "evidence of effectiveness"?

According to the *WWC* (and as outlined in *EDGAR* and *ESSA*), a program, policy, or practice (or intervention[17]) "demonstrates 'effectiveness' if the research has shown that it *caused* an improvement in outcomes."[18] Two common types of such research studies are efficacy studies, which look at the performance of an intervention under ideal and controlled circumstances (such as a program developer might do when creating a packaged intervention to improve children's reading), and effectiveness studies, which investigate the extent to which a particular intervention works in real-world settings (such as a school district might do when evaluating how well the packaged reading intervention can be implemented in the classroom).[19] These studies aim to *isolate* the effects of a particular intervention to determine whether and to what degree the intervention is *responsible* for the observed results.

---

[14] However, it's important to note that you will not be able to draw conclusions about whether the observed changes are the result of participation in the project activities. At most, you'll be able to make an association between project participation and the observed results.

[15] See the WWC Standards for criteria for judging the quality of single case designs.

[16] Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Belmont, CA: Wadsworth.

[17] Here we refer to interventions in a broad sense, meaning any program, policy, practice or combination of them that are intended to produce a particular outcome.

[18] WWC, 2015. Source: *http://ies.ed.gov/ncee/wwc/sitesearch.aspx?Search=effectiveness+study&website=NCEE%2fWWC&x=6&y=9*

[19] See for example, Singal, A., Higgins, P., & Waljee, A. (2014). A primer on effectiveness and efficacy trials. *Clinical and Translational Gastroenterology,* 45(5), e45, doi:10.1038/ctg.2013.13.

So, if you want to *empirically* demonstrate the **effectiveness** of your project—as defined by the WWC—you'll need to include in your project plan a rigorous evaluation that features an experimental design, a quasi-experimental well-matched control-group design, a regression discontinuity design, or a single-case design study that allows for demonstrations of effectiveness. However, as outlined in the tables below and on the next page, there are a number of factors that might hinder an OSEP project's ability to show "evidence of effectiveness" as defined by the WWC.

Despite these factors, it's still possible for you to conduct an evaluation that shows whether and to what degree your project has achieved its expected outcomes, as discussed in the next sections.

How can I know if the project is (or will be) successful?

The most straightforward way to know whether your project is **successful** is to design and conduct a high-quality evaluation that collects data on all three levels of outcomes (short-term, intermediate, and long-term) included in your logic model, and that includes a research design that would meet the WWC standards for demonstrating effectiveness. However, as mentioned above, evaluation designs such as this can be difficult to implement for a variety of reasons.

Study Design
Issues That May Affect an OSEP Project's Ability to Empirically Demonstrate "Effectiveness"

Potential Issue	Implication for the Evaluation
Insufficient time to fully capture intervention-associated outcomes	➤ There may be insufficient time available within the project's funded grant period to fully and comprehensively measure all possible outcomes. This may be due to the amount of time needed to fully implement a project as well as the amount of time for ultimate outcomes to complete their sequencing, mature, and become assessable.
Resource constraints that limit the scope of the evaluation	➤ The evaluation may not have the fiscal resources necessary use experimental or high-quality quasi-experimental designs to demonstrate a project's impact. These types of studies require sufficient samples of both treatment and control groups—in other words, populations that do and do not receive the project's services. A rigorous study also will require attention to the confounding factors identified in the theory of change. Altogether, resource constraints may limit the amount and types of data actually collected during an evaluation.
Ethical challenges to assigning individuals with disabilities to treatment and control or comparison groups	➤ Especially if a project is expected to have very beneficial results, stakeholders may not permit a study to assign individuals to treatment groups while at the same time withholding that same intervention (or providing a less effective intervention, or the *status quo*) from other individuals. In these cases, wait-list designs may be used, but they, too, are not always feasible.

Study Population
Issues That May Affect an OSEP Project's Ability to Empirically Demonstrate "Effectiveness"

Potential Issue	Implication for the Evaluation
The low frequency of certain disabilities in the study population *(Note: This issue also relates to study design because it can affect sample size and power to detect effects)*	➤ It may not be possible to find enough individuals with certain disabilities to include both treatment and control or comparison groups. ➤ It may not be possible to find individuals who would be considered good candidates for inclusion in a matched comparison group. ➤ It may not be possible for evaluators to include individuals with low-incidence disabilities in a study when reporting data may inadvertently identify those individuals. ➤ It might be difficult to find a large enough sample to meet the assumptions of certain statistical analyses.
Differences among individuals who share the same disability	➤ Even among individuals who share the same disability, it may be difficult for evaluators to demonstrate a good match between treatment and comparison groups on key characteristics.
Lack of assessments or tools that adequately capture outcomes for individuals with disabilities	➤ Many standardized assessments are developed and tested with the general population in mind. This is not always appropriate for assessing progress in exceptional children's populations. Tindal et al. (2010), for example, document the challenges of assessing children with special learning needs; issues include not only construction of the assessment but the modes of administration.[20] ➤ Many students with disabilities are given individualized accommodations during testing, making it difficult for evaluators to compare the results of the tests across individuals. ➤ Some students with disabilities are not included in testing.
The need to give consent to participate in a study	➤ Institutional Review Boards generally require consent from study participants, which may pose a problem if a particular individual is unable or a parent or guardian is unwilling to give consent.

Despite these issues, there are ways you can use evidence to show that your project is achieving its objectives. As discussed in the *Planning Phase*, if you made careful use of existing evidence when you created your theory of change and corresponding logic model, you've already made a good argument that, if carried out faithfully according to plan, your project will likely result in the eventual achievement of the long-term outcomes (e.g., improved outcomes for students with disabilities). When combined with high-quality *formative data* that show that participation in project activities has been high and the correct content has been covered (i.e., there is high

fidelity)—and especially if you can show that project participants haven't received similar training or services from another project or source—you can make a strong argument that the project has been responsible for the observed changes.[21] The example on the next page returns to the hypothetical Parent Resource and TA Center conducting trainings on autism for preschool lead teachers and shows how you might use data to demonstrate results—even without a comparison group.

---

[20] Tindal, G., Yovanoff, P., and Geller, J.P. (2010). Generalizability theory applied to reading assessments for students with significant cognitive disabilities. *The Journal of Special Education,* 44(1), 3-17.

[21] Reynolds outlines a method for making causal inferences based on strong theory and use of quantitative analytical techniques. Called confirmatory program evaluation, this method examines the pattern of empirical findings against several causal criteria, giving emphasis to identifying causal mechanisms or active ingredients of a program that yield effects. This method is most useful when there is extensive longitudinal data available and an established program theory. Source: Reynolds, A. J. (1998). Confirmatory program evaluation: A method for strengthening causal inference. *American Journal of Evaluation,* 19(2), 203-221.

What do I do if I can't collect extensive summative data?

At times, your project won't have the time or resources available to conduct a large-scale summative evaluation. In cases like this, a few strategies may be helpful.

- **Use available evidence to your advantage.** If you're using a research-based intervention, you can be reasonably sure that if you achieve high fidelity you can make inferences about expected project success. In these cases, invest project resources in collecting high quality progress monitoring, formative, and initial summative data to infer the likelihood of the project's eventual success.

- **Seek data sharing opportunities.** Partnerships with district and state education authorities may permit data sharing, or the use of district and state-collected data (such as end-of-grade tests), for project evaluation purposes. Data sharing allows evaluators to improve the rigor of the research design through strategies such as creating a matched comparison group while keeping the costs for evaluation relatively low. [22]

- **Narrow the focus of your summative evaluation.** It's possible to conduct smaller-scale, focused, studies that contribute important and informative evidence about your project. These may include experimental studies with limited range, in which a small set of very focused questions are addressed. This allows for a smaller sample size and narrower range of data collection while at the same time producing data that can be subject to powerful inferential statistical tests. Other options include well-designed single-case studies, many of which forego the inclusion of separate comparison or control groups. (Note that guidelines and organizing concepts for these studies are available through *What Works Clearinghouse*.)

Example: Using Data to Assess Project Results

Here we continue with our hypothetical Parent Resource and TA Center example.

Data sources
Training agendas, interviews with project staff, post-training teacher survey, online teacher survey (surveys have > 80% response rates), coaching logs, interviews with teachers getting coaching, fidelity data.

Evaluation design
Pre-post assessments of treatment group knowledge and skills (no comparison group).

Findings
➤ Short-term outcome achieved—Summative data shows improvements in preschool lead teachers' ability to recognize the early signs of autism among young children.

➤ Intermediate outcome achieved—Summative data shows teachers are able to apply their skills of identification of and support for children in their practice.

➤ Fidelity data show that the project has been implemented with fidelity to the model.

Ruling out possible alternate explanations (or confounds)
➤ District staff interviews and training agendas show that no content related to recognizing the early signs of autism had been included in the district's professional development.

➤ Teacher survey indicates that teachers have not received training in this topic from other organizations.

Conclusions
Even without a comparison group, the evaluators feel confident that the project was responsible for the changes in participating teachers' knowledge and skills in this area.

➤ Based on the logic model, given that both the short-term and intermediate outcomes have been achieved—and knowing that the project has been implemented with fidelity—it can be reasonably concluded that the project, if sustained, is likely to achieve its long-term outcome of improving the early identification of and support for children with autism and their families.

➤ The project has established several implementation strategies (e.g., trained coaches) that should help to sustain the practice after the project's exit.

---

[22] States and districts have data sharing protocols, including the use of informed written consent, which will need to be adhered to in order to access the desired data.

How can I make projections about whether the project will ultimately achieve its long-term outcomes?

As mentioned previously, many OSEP projects will not be able to determine whether the long-term outcomes are ultimately achieved. However, you might be able to make inferences about the project's long-term, summative, success through techniques such as outcome sequencing and forecasting. **Outcome sequencing** suggests that outcome achievement occurs in predictable ways, starting with an immediate reaction to the intervention (such as training or professional development), followed by a secondary effect, which is learning (for an example, see the *Kirkpatrick Model*). This is then followed by a tertiary change, a shift in behavior, and then results, which are the changes in participation that connect back to the originating need. To apply outcome sequencing in practice, you can identify short-term and intermediate levels of change that are predictive of the longer-term summative outcomes the project is hoping to influence. Tying all of these concepts together, you can use available project and administrative data to project the expected summative outcomes, based on documentation of short-term and intermediate outcomes and considering level of fidelity and any barriers to or facilitators of project implementation.

Forecasting techniques rely upon existing data as well as a project's research base. Forecasting techniques require data collected over multiple periods of time that can be used to make predictions about future trends, either as a point value (a single, numerical value such as a test score) or as an interval (a range of scores in which the actual value may lie). A project team might use available data from other projects or state assessments to make predictions about how student achievement scores may change after an intervention is implemented. Forecasting techniques allow evaluators to project possible values,

with a certain amount of confidence. Sensitivity analyses associated with forecasting help establish the degree of confidence evaluators can have in the forecasting model. As a general rule of thumb, the better the data available to construct the forecast, the more accurate the model. It's important to consider that projects that plan to use forecasting may need to (a) collect more short-term and intermediate data than otherwise would be necessary, and (b) use the existing research base to construct more than one possible model of long-term outcomes. We suggest you consult a statistician for more information.

How can I use data to inform decisions related to project continuation, replication, or dissemination?

Once you've gone through the process of gathering and using evidence to monitor progress, assess implementation and provide formative feedback, and determine whether your project has achieved its expected outcomes, you'll likely have good data to inform decisions related to project continuation, replication, or dissemination. Project staff who have data that show that their project has been implemented with fidelity and achieved its objectives can then use that data to market their projects, including applying for additional or new resources to support project continuation, replication, or dissemination. For example, such data can be used to:

- Confirm that the model, intervention, or services implemented are effective and efficient in achieving desired and meaningful outcomes;
- Confirm that the project team can successfully deploy projects of similar scope;
- Establish the need to replicate or scale-up the project into new locations or populations; and
- Make adjustments prior to replication of the project in different settings or dissemination of findings from the project.

Summary

There are myriad ways evidence can be used across the project cycle. As we've discussed, evidence can be used to identify and support theories of action and choice of interventions, guide logic model development and evaluation planning, provide information on implementation progress and fidelity, and demonstrate project results. For each phase of the project cycle, we identify types of evidence that might be collected, where that evidence might be found, and how it might be produced and used. While it might not always be possible to gather or produce the best available research evidence

on your project, it's important to remember that if you make careful use of existing evidence when planning your project, gather high-quality formative evaluation data during the implementation phase, and collect summative evaluation data on short-term and intermediate outcomes during the results phase, you can make a good argument that your project will likely produce the expected results. And, by doing so, you're contributing evidence that may support continuing your ongoing project activities, disseminating your findings, and even replicating or scaling-up your project.

Appendix

EDGAR Regulations Related to Demonstrating Evidence

34CFR 77.1

http://www.ecfr.gov/cgi-bin/text-idx?SID=393301a7cdccca1ea71f18aae51824e7&node=34:1.1.1.1.24&rgn=div5

§77.1 Definitions that apply to all Department programs.

EDGAR Evidence of Promise Definition: empirical evidence to support the theoretical linkage(s) between at least one critical component and at least one relevant outcome presented in the logic model for the proposed process, product, strategy, or practice. Specifically, evidence of promise means the conditions in both paragraphs (i) and (ii) of this definition are met:

(i) There is at least one study that is a—

(A) Correlational study with statistical controls for selection bias;

(B) Quasi-experimental design study that meets the What Works Clearinghouse Evidence Standards with reservations; or

(C) Randomized controlled trial that meets the What Works Clearinghouse Evidence Standards with or without reservations.

(ii) The study referenced in paragraph (i) of this definition found a statistically significant or substantively important (defined as a difference of 0.25 standard deviations or larger) favorable association between at least one critical component and one relevant outcome presented in the logic model for the proposed process, product, strategy, or practice (EDGAR 77.1, Definitions).

Quasi-experimental design study means a study using a design that attempts to approximate an experimental design by identifying a comparison group that is similar to the treatment group in important respects. These studies, depending on design and implementation, can meet What Works Clearinghouse Evidence Standards with reservations (but not What Works Clearinghouse Evidence Standards without reservations).

Randomized controlled trial means a study that employs random assignment of, for example, students, teachers, classrooms, schools, or districts to receive the intervention being evaluated (the treatment group) or not to receive the intervention (the control group). The estimated effectiveness of the intervention is the difference between the average outcomes for the treatment group and for the control group. These studies, depending on design and implementation, can meet What Works Clearinghouse Evidence Standards without reservations.

Moderate evidence of effectiveness means one of the following conditions is met:
(i) There is at least one study of the effectiveness of the process, product, strategy, or practice being proposed that meets the What Works Clearinghouse Evidence Standards without reservations, found a statistically significant favorable impact on a relevant outcome (with no statistically significant and overriding unfavorable impacts on that outcome for relevant populations in the study or in other studies of the intervention reviewed by and reported on by the What Works Clearinghouse), and includes a sample that overlaps with the populations or settings proposed to receive the process, product, strategy, or practice.

(ii) There is at least one study of the effectiveness of the process, product, strategy, or practice being proposed that meets the What Works Clearinghouse Evidence Standards with reservations, found a statistically significant favorable impact on a relevant outcome (with no statistically significant and overriding unfavorable impacts on that outcome for relevant populations in the study or in other studies of the intervention reviewed by and reported on by the What Works Clearinghouse), includes a sample that overlaps with the populations or settings proposed to receive the process, product, strategy, or practice, and includes a large sample and a multi-site sample.

Note: Multiple studies can cumulatively meet the large and multi-site sample requirements as long as each study meets the other requirements in this paragraph.

Strong evidence of effectiveness means one of the following conditions is met:
(i) There is at least one study of the effectiveness of the process, product, strategy, or practice being proposed that meets the What Works Clearinghouse Evidence Standards without reservations, found a statistically significant favorable impact on a relevant outcome (with no statistically significant and overriding unfavorable impacts on that outcome for relevant populations in the study or in other studies of the intervention reviewed by and reported on by the What Works Clearinghouse), includes a sample that overlaps with the populations and settings proposed to receive the process, product, strategy, or practice, and includes a large sample and a multi-site sample.

Note: Multiple studies can cumulatively meet the large and multi-site sample requirements as long as each study meets the other requirements in this paragraph.

(ii) There are at least two studies of the effectiveness of the process, product, strategy, or practice being proposed, each of which: Meets the What Works Clearinghouse Evidence Standards with reservations, found a statistically significant favorable impact on a relevant outcome (with no statistically significant and overriding unfavorable impacts on that outcome for relevant populations in the studies or in other studies of the intervention reviewed by and reported on by the What Works Clearinghouse), includes a sample that overlaps with the populations and settings proposed to receive the process, product, strategy, or practice, and includes a large sample and a multi-site sample.

Every Student Succeeds Act (ESSA) (Public Law 114–95) Language Related to Demonstrating Evidence

http://edworkforce.house.gov/uploadedfiles/every_student_succeeds_act_-_conference_report.pdf

'(21) EVIDENCE-BASED.—

"(A) IN GENERAL.—Except as provided in subparagraph (B), the term 'evidence-based', when used with respect to a State, local educational agency, or school activity, means an activity, strategy, or intervention that—

"(i) demonstrates a statistically significant effect on improving student outcomes or other relevant outcomes based on—

"(I) strong evidence from at least 1 well-designed and well-implemented experimental study;

"(II) moderate evidence from at least 1 well-designed and well-implemented quasi-experimental study; or

"(III) promising evidence from at least 1 well-designed and well-implemented correlational study with statistical controls for selection bias; or

"(ii)(I) demonstrates a rationale based on high-quality research findings or positive evaluation that such activity, strategy, or intervention is likely to improve student outcomes or other relevant outcomes; and

"(II) includes ongoing efforts to examine the effects of such activity, strategy, or intervention.

"(B) DEFINITION FOR SPECIFIC ACTIVITIES FUNDED UNDER THIS ACT.—When used with respect to interventions or improvement activities or strategies funded under section 1003, the term 'evidence-based' means a State, local educational agency, or school activity, strategy, or intervention that meets the requirements of subclause (I), (II), or (III) of subparagraph (A)(i)."

Summary of What Works Clearinghouse Effectiveness and Extent of Evidence Ratings[23]

The WWC effectiveness ratings range from negative effects to positive effects, as outlined below:

- **Positive effects:** strong evidence that intervention had a positive effect on outcomes.
- **Potentially positive effects:** evidence that intervention had a positive effect on outcomes with no overriding contrary evidence.
- **Mixed effects:** evidence that intervention's effect on outcomes is inconsistent.
- **No discernible effects:** no evidence that intervention had an effect on outcomes.
- **Potentially negative effects:** evidence that intervention had a negative effect on outcomes with no overriding contrary evidence.
- **Negative effects:** strong evidence that intervention had a negative effect on outcomes.

For an individual study of a program, policy or practice (referred to as "interventions"), the WWC rating of effectiveness takes into account four factors:

- the quality of the research on the intervention (as assessed using the *WWC standards*);
- the statistical significance of the research findings;
- the size of the differences between participants in the intervention and comparison groups; and
- the consistency in findings across studies.

Intervention reports are categorized according to the extent of evidence that supports the findings. The extent of evidence categorizations focus on the number and sizes of studies of the intervention in order to give an indication of how broadly findings may be applied to different settings. There are two extent of evidence categories:

- **Small:** includes only one study, or one school, or findings based on a total sample size of less than 350 students and 14 classrooms (assuming 25 students in a class);
- **Medium to large:** includes more than one study, more than one school, and findings based on a total sample of at least 350 students or 14 classrooms.

---

[23] See the WWC Standards 3.0 Handbook for more information: *http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf*

OSEP Logic Model Outline

The OSEP Logic Model Outline presented below illustrates the different elements that are commonly included in logic models. It also presents information to guide completion of the logic model, for example, by showing that short-term outcomes are often focused on changes in learning, intermediate (or medium-term) outcomes often focus on changes in action, and long-term outcome focus on changes in conditions.
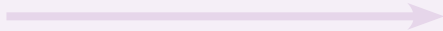
OSEP Logic Model Outline

Inputs	Activities	Outputs	Outcomes-Impact		
			Short-term	Medium-term	Long-term
What we invest	**What we do**	**What we produce**	**What the short-term results are**	**What the medium-term results are**	**What the long-term results are**
Staff	Develop/ implement programs	# of activities conducted	Learning	Action	Conditions
Volunteers	Develop/ implement direct services	Quality of activities conducted	Awareness	Behavior	Social
Time	Develop products, technology, curriculum, resources	Fidelity of activities conducted	Knowledge	Practice	Economic
Money		Number of targeted clients reached	Attitudes	Decision making	Civic
Research base	Train	Number of participants	Skills	Policies	Environmental
Materials	Research	Characteristics of participants	Opinions	Social action	
Equipment	Assess		Aspirations	Organizational change	
Technology	Facilitate		Motivations		
Partners	Partner/ collaborate				
	Disseminate				

External Factors

Evaluation: Focus ✦ Collect Data ✦ Analyze ✦ Interpret ✦ Report

Logic Model Example

Sample Logic Model for a Hypothetical Parent Resource and Technical Assistance Center

Outcomes →

Inputs	Activities	Outputs	Short-term	Medium-term	Long-term
➤ Staff with appropriate knowledge and expertise ➤ Funding ➤ Community agencies	➤ Develop and provide resources and materials to parents of children with disabilities ➤ Develop and provide training workshops to parents of children with disabilities ➤ Establish and maintain collaborative relationships with external agencies	➤ Number and types of resources materials developed ➤ Materials distributed ➤ Number of training workshops ➤ Content of training workshops ➤ Number of parents participating in training workshops ➤ Characteristics of participants	➤ Parents have increased knowledge of strategies to support and encourage their children's development and learning ➤ Parents have increased awareness of relevant supports and services available to them	➤ Parents feel more confident and competent in supporting their child's development and learning ➤ Parents access support services and activities when needed	➤ Parents are better able to offer support to children's learning and development ➤ Children show improvement in learning and development

Sample Section of a Fidelity Matrix

In the section that discusses *how to know if the project is being implemented with fidelity*, we outlined five steps for creating a fidelity measurement system:

1. **Identify the "key components" of the project, or those features that are critical for the project to achieve positive results.** These key components should be clearly illustrated in the project logic model.

2. **Operationally define each key component (e.g., professional development) included in the logic model.** An operational definition is composed of indicators, which are specific aspects of the intervention that can be measured quantitatively. The number of indicators should reflect the complexity of the project. It is best to have more than one indicator of a specific component, preferably from different sources. Additionally, it's best if the indicators are able to differentiate among different levels of fidelity. Be sure to identify indicators that can actually be measured.

3. **Select data sources and measures.** To do this, consider the research questions guiding the study, find the best data sources for each indicator, and choose whether to use existing instruments or develop new ones. Use multiple sources of data for each indicator (e.g., surveys and observations) when possible, and consider reliability/validity of data sources and measures.

4. **Establish fidelity thresholds.** Fidelity thresholds tell you the least amount of each indicator (or component) that need to be present for fidelity to be considered "adequate." Represented as numeric scales, thresholds quantify the extent to which an indicator was enacted with fidelity. This scale can be dichotomous (0 or 1) or it can range from 0 to 3, or even 0 to 5.

5. **Calculate fidelity scores.** Fidelity scores are calculated based on the fidelity thresholds and tell you the least amount of each component that needs to be present for fidelity to be considered "adequate." You may want to calculate a fidelity score for each key component separately, or create one score for fidelity across all project components.

The table on the next page presents an example of one section of a fidelity matrix that could be used for the Parent Resource and TA Center example. In the column headers, the numbers in parentheses refer to the steps outlined above.

1. Key Component of Training	2. Operational Definition & Indicators	3. Data Sources/ Measures	4. Fidelity Thresholds & 5. Score for Adequate Fidelity
Deliver training to families of children with autism	**Operational Definition:** To be considered implemented with fidelity, (a) the trainings will be delivered by lead parent educators who have the required competencies, (b) training workshops will be delivered on schedule and within the allotted timeframe, and will cover the expected content; and (c) parents will participate in and express satisfaction with the workshops. **Indicator A:** Staff competencies. ➤ Lead parent educators who will provide training to parents have: i. a Bachelor's Degree or an Associate's Degree in Special Education, Early Childhood Education, or a related field; ii. at least one year of experience working with families of children with disabilities; and iii. positive evaluations that provide evidence that the staff member understands the content being presented, effectively interacts with families, and effectively facilitates training workshops.	➤ Administrative data on staff credentials and experience ➤ Staff evaluations	**Indicator A-i:** Education (program-level threshold) **0 =** > 75% of lead parent educators on staff meet competency indicator **1=** 75%-89% of lead parent educators on staff meet competency indicator **2 =** ≥ 90% of lead parent educators on staff meet competency indicator **Indicator A-ii:** Experience with families (program-level threshold) **0 =** > 75% of lead parent educators on staff meet competency indicator **1=** 75%-89% of lead parent educators on staff meet competency indicator **2 =** ≥ 90% of lead parent educators on staff meet competency indicator **Indicator A-iii:** Positive evaluations (program-level threshold) **0 =** > 60% of lead parent educators on staff meet competency indicator **1=** 60%-75% of lead parent educators on staff meet competency indicator **2 =** ≥ 75% of lead parent educators on staff meet competency indicator **Score to achieve "adequate fidelity" = 4** (Note: To achieve adequate fidelity the score for education must = 2)

1. Key Component of Training	2. Operational Definition & Indicators	3. Data Sources/ Measures	4. Fidelity Thresholds & 5. Score for Adequate Fidelity
Deliver training to families of children with autism	**Indicator B:** Delivery of Training Workshops: i. 4 training workshops are delivered on schedule (within allotted timeframe) ii. Criteria for content and format of training workshops: ➤ addresses the expressed needs of parents ➤ encourages parents to participate and share their experiences, and ➤ incorporates at least three training methods (e.g., lecture, demonstration, technology, games, skill practice, group discussion).	➤ Needs assessment data ➤ Workshop schedules, agendas and materials ➤ Observation rubric or checklist ➤ Participant feedback surveys	**Indicator B-i:** Delivery of training (program-level threshold) **0 =** 0-1 workshops are delivered within allotted timeframe **1=** 2-3 workshops are delivered within allotted timeframe **2 =** 4 workshops are delivered within allotted timeframe **Indicator B-ii:** Content and format of training (program-level threshold) **0 =** 0-1 workshops meet the criteria for the content and format of the training **1 =** 2-3 workshops meet the criteria for the content and format of the training **2 =** All workshops meet the criteria for the content and format of the training ———— **OR** ———— **0 =** None of the criteria for content and format are met for the training workshops **1 =** 2 of 3 criteria for content and format are met for the training workshops **2 =** All 3 criteria for content and format are met for the training workshops **Score to achieve "adequate fidelity" = 3** [Note: To achieve adequate fidelity the score for delivery of training must = 2]

1. Key Component of Training	2. Operational Definition & Indicators	3. Data Sources/ Measures	4. Fidelity Thresholds & 5. Score for Adequate Fidelity
Deliver training to families of children with autism	**Indicator C:** Parent Participation in and Satisfaction With Training Workshops: i. Percent of parents that attend the workshops ii. Percent of parents in attendance that: ➤ actively participate in discussions, share experiences, or ask questions; ➤ complete all activities; and ➤ are satisfied with the quality, relevance and usefulness of the workshops.	➤ Workshop attendance records ➤ Observation rubric or checklist ➤ Participant feedback surveys	**Indicator C-i:** Parent attendance (program-level threshold) **0 =** > 50% of parents attend the workshops **1 =** 51%-75% of parents attend the workshops **2 =** > 75% of parents attend the workshops **Indicator C-ii:** Parent participation and satisfaction (program-level threshold) **0 =** > 75% of parents who attend meet all three participation and satisfaction criteria **1 =** 75%-89% of parents who attend meet all three participation and satisfaction criteria **2 =** ≥ 90% of parents who attend meet all three participation and satisfaction criteria ——— **OR** ——— **0 =** Parents who attend workshops meet none of the criteria for participation and satisfaction **1 =** Parents who attend workshops meet 2 of the 3 criteria for participation and satisfaction **2 =** Parents who attend workshops meet all 3 of the criteria for participation and satisfaction ——— **OR** ——— **0 =** > 75% of parents who attend actively participate and complete all activities **1 =** 75%-89% of parents who attend actively participate and complete all activities **2 =** ≥ 90% of parents who attend actively participate and complete all activities ——— **AND** ——— **0 =** > 75% of parents who attend are satisfied with the quality, relevance and usefulness of the workshops **1 =** 75%-89% of parents who attend are satisfied with the quality, relevance and usefulness of the workshops **2 =** ≥ 90% of parents who attend are satisfied with the quality, relevance and usefulness of the workshops **Score to achieve "adequate fidelity" =** Depends on the selection of thresholds