Institute of Education Sciences

Conducting implementation research in impact studies of education interventions:
A guide for researchers

A Publication of the National Center for Education Evaluation and Regional Assistance

U.S. Department of Education
Miguel Cardona
Secretary

Institute of Education Sciences
Mark Schneider
Director

National Center for Education Evaluation and Regional Assistance
Matthew Soldner
Commissioner

Thomas Wei
Amy Johnson
Project Officers

The Institute of Education Sciences (IES) is the independent, non-partisan statistics, research, and evaluation arm of the U.S. Department of Education. The IES mission is to provide scientific evidence on which to ground education practice and policy and to share this information in formats that are useful and accessible to educators, parents, policymakers, researchers, and the public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other IES product or report, we would like to hear from you. Please direct your comments to ncee.feedback@ed.gov.

This report was prepared for the Institute of Education Sciences (IES) under Contract 91990020F0052 by Mathematica. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

June 2023

Hill, C. J., Scher, L., Haimson, J., & Granito, K. (2023). Conducting implementation research in impact studies of education interventions: A guide for researchers. (NCEE 2023-005). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee.

This report is available on the Institute of Education Sciences website at http://ies.ed.gov/ncee.

![Institute of Education Sciences logo - IES]

Conducting implementation research in impact studies of education interventions: A guide for researchers

Carolyn J. Hill (MDRC)
Lauren Scher (Mathematica)
Joshua Haimson (Mathematica)
Kelly Granito (MDRC)

Contents

Exhibits

Conducting implementation research in impact studies of education interventions: A guide for researchers

Implementation analyses conducted as part of impact studies can help educators know whether a tested intervention is likely to be a good fit for their own settings. This guide can help researchers design and conduct these kinds of analyses. The guide provides steps and recommendations about ways to specify implementation research questions, assess whether and how the planned intervention is implemented, document the context in which the intervention is implemented, and measure the difference between the intervention and what members of the control group receive. It presents strategies for analysis and reporting about these topics, and also for linking implementation and impact findings. The guide offers key definitions, examples, templates, and links to resources.

Introduction

Evaluations of education interventions can help decision makers improve schools and programs in ways that help all learners succeed and achieve their full potential. High-quality impact studies of education interventions provide the strongest evidence about whether interventions improve academic outcomes such as school readiness, achievement, learning, persistence, or graduation; social and behavioral competencies; or employment and earnings outcomes. Yet information about whether and by how much a tested intervention improves outcomes is only part of the story. To learn why and how impact findings vary and to support the broader use of effective interventions, educators need to understand how, and under what conditions, the intervention was implemented. High-quality implementation research can contribute to these understandings.

Why this guide

This guide provides recommendations and specific steps to help researchers conducting impact studies develop, plan, and report findings from implementation research to further contribute to the evidence base to improve student outcomes. This guide, sponsored by the U.S. Department of Education's Institute of Education Sciences (IES), is one in a series of guides designed to strengthen education research. Through its Standards for Excellence in Education Research (SEER; IES, 2022a), IES seeks to ensure that evidence-building involves deep and nuanced understanding about the tested interventions and their implementation, context, and cost.

SEER "codifies practices that IES expects–and increasingly requires–to be implemented as part of IES-funded causal impact studies" (IES, 2022a). Causal impact studies are one type of

study within the broader field of program evaluation, which also includes evaluability assessments, process studies, cost studies, and monitoring studies (Rossi et al., 2018).

Similarly, the type of implementation research described in this guide is one type of study within the broader field of implementation research. The broader field addresses whether and how policies, practices, procedures, and interventions are put into use by frontline staff, managers, policymakers, and other actors. It examines strategies that actors use, contexts in which implementation occurs, and facilitators and barriers to implementation. The broader field includes studies conducted as part of evaluations, as well as studies that are not evaluations but that aim to describe the implementation of interventions, programs, or policies; test hypotheses about implementation of these interventions, programs, or policies; or improve practice.

The current guide focuses on a particular type of implementation research for a particular form of program evaluation: **conducting implementation research as part of high-quality impact studies of education interventions**.[1] This focus aims to support researchers in applying the SEER standard about documenting treatment implementation and contrast (Box 1).[2]

The guide includes recommendations for researchers planning high-quality impact studies on how they might study intervention implementation. It emphasizes the critical pieces of information implementers need to interpret findings from an impact study, including how and in what contexts a researcher's intervention is implemented and how the intervention compares with other available programs or interventions in the study's settings. The guide encourages systematic planning for implementation research, including collecting, analyzing, and reporting information about

Box 1. Standards for excellence in education research

Document treatment implementation and contrast

1. Researchers must document how, and the context within which, the treatment was implemented.

2. Researchers must document the counterfactual condition, including its context.

3. Researchers must measure the essential elements of the treatment contrast between the treatment and control conditions.

4. Researchers must measure the fidelity of an intervention's implementation.

Recommendation

1. Researchers should document, and identify opportunities to learn from, adaptations of the intervention that were observed during implementation.

Source: IES (2021a).

---

[1] The implementation science literature refers to the kind of implementation research discussed in this guide as a Hybrid Type I or Hybrid Type II design (Brown et al., 2017; Curran et al., 2012). The guide does not mention these terms elsewhere.

[2] The SEER standards use the term "treatment" to refer to the tested intervention. "Intervention" and "program" are terms used by researchers, funding agencies, or publications to refer to the same construct as "treatment" in a program evaluation context. This guide uses the term "intervention."

implementation and context. In so doing, the guide seeks to ensure that education research investments contribute rich understanding about how to improve student outcomes.

Who should use this guide

The primary audience for the guide is researchers who conduct or direct experimental impact studies–primarily randomized controlled trials. These researchers include applicants for, and recipients of, research funding from IES and other funders. Because some researchers may not have expertise in implementation research, the guide seeks to build their knowledge and capacity about implementation research by providing accessible descriptions and references to additional resources. The guide is a starting point to this capacity building: studies are stronger when their team members reflect a range of expertise, including those with strong backgrounds in impact design and analysis, implementation research, cost research, and technical assistance.

The guide is relevant to three additional audiences. The first includes researchers who conduct impact studies either prospectively or retrospectively using quasi-experimental or non-experimental designs such as regression discontinuity, propensity score, or comparative interrupted time series. Although the guide focuses on experimental impact studies that allow for prospective design, information in the guide about specifying the intervention, its implementation, its context, and its contrast are all ideas that can apply to nonexperimental impact design studies and to retrospective studies as well. The second audience includes graduate students with an interest in rigorous social science research. Although this guide is framed within the context of education and the education sciences, its lessons may be broadly applicable across the human services. Finally, research funders can use the guide's information to engage with researchers and support high-quality impact studies that include robust studies of intervention implementation.

Understanding why intervention effects vary

Information about a study's intervention, how it was implemented, the study context (including sample characteristics), and comparisons to the control condition can contribute to understanding how and why interventions are effective.[3]

Conceptual framework

A conceptual framework is useful to describe how implementation analyses contribute to impact studies. The framework in this guide, adapted from Weiss et al. (2014), traces the path

---

[3] Other phrases used include "getting inside the black box," "unpacking the logic model," and "examining variation (or heterogeneity) in intervention (or program or treatment) effects." As mentioned in Footnote 2, this guide uses the term "intervention" instead of "program" or "treatment."

from a planned intervention to impacts on student, teacher, or school outcomes (Exhibit 1). Much of the remaining guide focuses on three parts of the framework:

1. **Intervention model** (see upper left area of Exhibit 1). The intervention model includes both **direct components** and **support components**, with three broad stages for each type of component: **planned**, **offered**, and **received**. Section II discusses the intervention model in further detail.

   Direct components involve direct interaction with the intended population[4] whose impacts are hypothesized to change as a result of the intervention and whose impacts you will measure. For example, direct components in an intervention for first-year college students that aims to increase course completion and graduation rates might be course enrollment, advising, tutoring, and financial assistance.

   Support components are the strategies, resources, or preparation that activate or support the direct components in the study and that are specified by the study team, developers, or implementers as part of the intervention model. Support components, as discussed in this guide, correspond with the "implementation plan" in Weiss et al. (2014) and with "implementation strategies" in the implementation science literature (for example, Cook et al., 2019; Lyon & Bruns, 2019; Powell et al., 2015).[5] Depending on the intervention, support components might include providing professional development, training, coaching, or peer learning; providing supervision, monitoring, or feedback; conducting continuous quality improvement processes; conducting outreach or sharing information; specifying job descriptions or requirements; or setting up and using record-keeping or data systems. A specific factor (such as dedicated space for tutoring sessions, or record-keeping systems for monitoring student progress) can be a support component in one study but a context feature (discussed below) in another study. It is considered a support component if it is pre-specified as part of the intervention by the intervention developer, study team, or implementer; and it is considered a context feature if not.

   It is helpful for planning and analysis purposes to distinguish components' three distinct stages, as shown in Exhibit 1: as **planned** by the study team, developers, or implementers; as **offered** by implementers, such as technical assistance providers or teachers, to the intended target(s) of the component; and as actually **received** by the intended target of the component. This distinction between offered and received may be more relevant for some interventions (such as the number of career advising office hours postsecondary

---

[4] In education studies, intended populations may be students, teachers, classrooms, administrators, schools, districts, or systems, depending on the intervention.

[5] General introductions to and resources about implementation science include Bauer & Kirchner (2020), Curran (2020), and the University of Washington's Implementation Science Resource Hub (https://impsciuw.org/, University of Washington, 2023a). The Research Institute for Implementation Science in Education (RIISE, https://smartcenter.uw.edu/programs-services/riise/, University of Washington, 2023b) at the University of Washington is an example of a center that builds on implementation science to support research about implementation strategies in school settings. Examples of earlier-generation implementation science publications that did not use the terminology of "strategies" include Damschroder et al. (2009), Dane & Schneider (1998), Durlak & DuPre (2008), and Fixsen et al. (2005).

students were offered versus actually showed up for), but less relevant for others (such as a teacher providing guided instruction to students who are present in the classroom, who simultaneously receive the offered intervention).

Intervention fidelity refers to the correspondence between the intervention components that were *planned* and the intervention components that were *offered:* greater intervention fidelity indicates greater correspondence between the planned and implemented intervention model. Intervention fidelity can be assessed for each of four pre-specified dimensions--content, quantity, mode, and quality--which largely capture adherence/mode, exposure, and quality/responsiveness aspects of program integrity described in intervention fidelity frameworks by Dane & Schneider (1998), Cordray & Pion (2006), Hulleman & Cordray (2009), among others.[6]

2. **Intervention contrast** (see middle area of Exhibit 1) refers to the difference in relevant services received by the intervention and control group members in the study. Section IV of this guide discusses intervention contrast in further detail.

   Other terms for the intervention contrast include service contrast, treatment contrast, service differential, program differentiation, and achieved relative intervention strength. Some intervention fidelity frameworks, such as those mentioned above, include intervention contrast as part of intervention fidelity or integrity. As shown in Exhibit 1, the guide specifies intervention contrast as distinct from intervention fidelity.

   Depending on the study objectives, the counterfactual condition–what is available to and received by control group members–may be as specific as a less intensive or different form of the intervention (sometimes referred to as an A/B test). Or it may be as general as what services study participants in the control group would otherwise have access to in the classroom, school, or community. Researchers often use the phrase "business as usual" to refer to the latter situation. The counterfactual concept in experimental and non-experimental impact studies is the same: in the absence of the intervention, what services would the intervention group have received?

   It is important to consider the intervention contrast early and often throughout a study (Hamilton & Scrivener, 2018). For example, assessing whether the intervention contrast is likely to be small from the outset of a study can inform the need for stronger implementation of intervention components (perhaps through technical assistance).

3. **Context features** (see bottom of Exhibit 1) include characteristics of the organizational or programmatic setting where the intervention model is being tested, characteristics of the intended recipients, and aspects of the external environment. Context features are relevant for both experimental and nonexperimental designs. The relevant context features will depend on the intervention's logic model, the study's research questions, and the specific study design (for example, the point of randomization in a randomized controlled trial, or the time period and locations over which data are collected and analyzed for a comparative interrupted time series design). In the implementation science

---

[6] Section II of this guide discusses these four **dimensions** in further detail, following Weiss et al. (2014).

Exhibit 1. A conceptual framework for linking the intervention model, contrast, and impacts
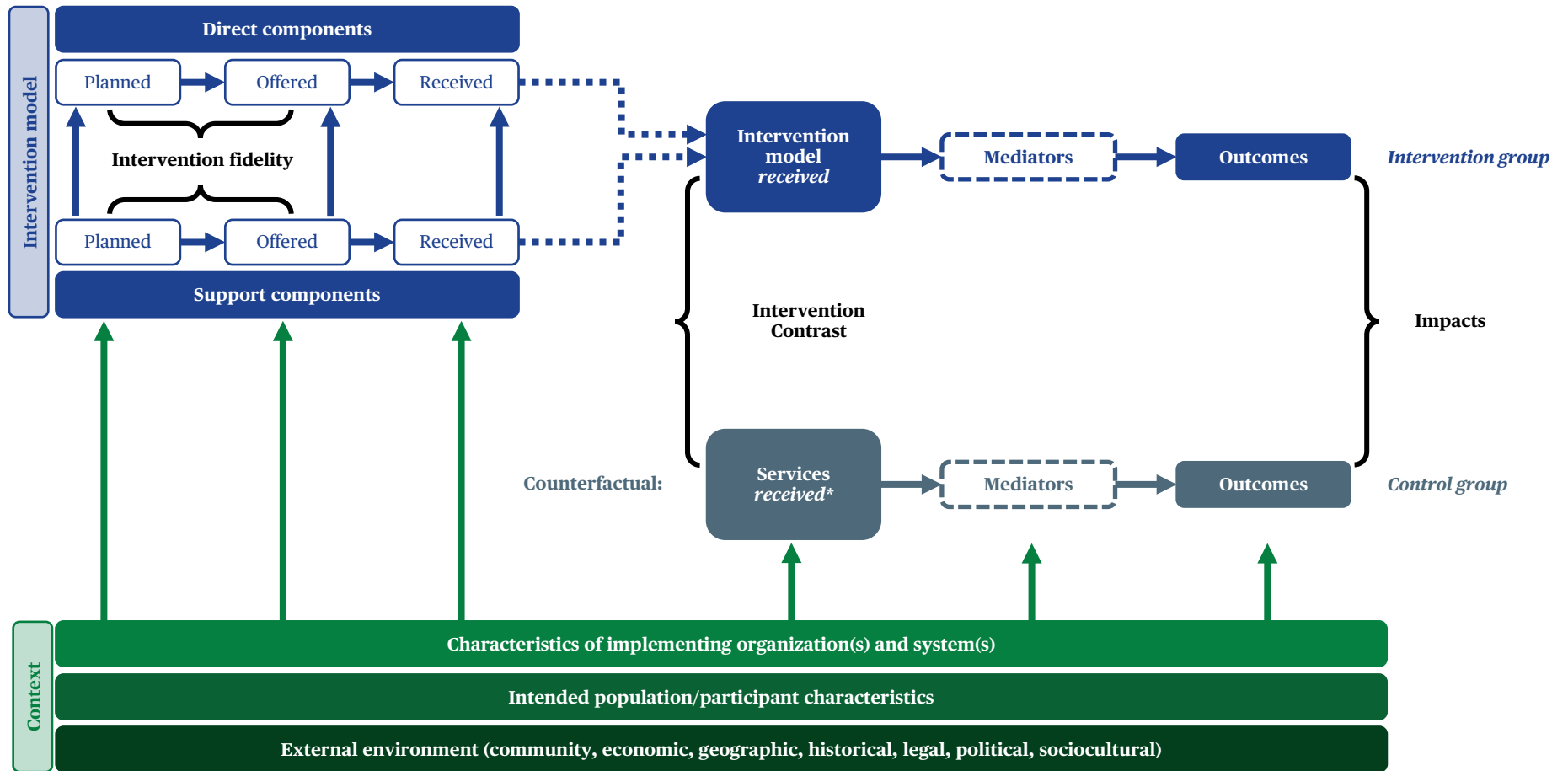
Source: Adapted from Weiss et al. (2014).

Note: Any services that are planned and offered to the control group in the counterfactual settings are not shown in this exhibit.

literature, "determinants" or "contextual determinants" are terms often used for context features; however, terms such as "barriers, hind[rance]s, obstacles, impediments, enablers, and facilitators" are also common (Nilsen & Bernhardsson, 2019, p. 3). Section III of the guide discusses context in further detail.

The relationship between intervention contrast and context depends on the specific context features considered, the intervention and its logic model (discussed in Section II), and the study design. For example, "business as usual" conditions often are influenced by what is required or allowed by school, districts, or state policies. Policy changes during the study may affect the intervention contrast if the change newly requires or allows changes in counterfactual conditions that are relevant for control group members (described in Section IV).

Research on the intervention model, intervention contrast, and context are typically part of implementation research conducted within impact studies of education interventions. Section V addresses how they moderate and mediate impacts.

Examples of how implementation research can inform impacts

Implementation research can help explain why an intervention might not generate effects (that is, "null findings"). Null findings might occur for a variety of reasons (Jacob et al., 2019). They might arise because an otherwise effective intervention was not implemented as intended (that is, poor implementation of the intervention, leading to no impacts); because the theory behind the intervention was not valid (that is, strong implementation of an otherwise ineffective intervention or an intervention designed for a different context or population, leading to no impacts); or because what was offered to and received by intervention and control group members was not sufficiently different to generate a detectable improvement of outcomes for the intervention group over and above the outcomes for the control group (that is, the intervention contrast was weak).

Evidence from implementation research can provide insight into why an intervention is more effective in some implementing sites than others. Even when the study design does not involve planned variation by site in interventions or in their implementation, the implementation research provides insights into possible reasons that impacts vary across sites. For example, a study of schools managed by charter management organizations found considerable variation in impacts across schools on middle school reading and math achievement (Furgeson et al., 2012).[7] Using information collected through its implementation research, the study found that these impacts were positively associated with schools' use of comprehensive behavior policies and with intensive teacher training.

---

[7] The study used propensity score matching to generate achievement impact estimates for the 22 schools and replicated the impact estimates using random assignment for a subset of the schools.

Findings from implementation research in a single study can provide explanations for why some specific interventions, or specific components, might be more effective than others. Some aspects of the intervention might vary naturally, or the interventions may be randomly assigned. For example, one study examined the effectiveness of early math curricula by randomly assigning schools to one of four curricula, then measuring student achievement (Agodini et al., 2010).[8] The impact analyses found differences in student math achievement across the four curricula. It was important to establish that each of the four curricula were implemented as intended, and that the impact study offered a fair test of the curricula. The study collected implementation data through surveys and classroom observations, which indicated that teachers used the curriculum they were assigned and that their instructional practices, such as student-centered or teacher-centered instruction, were consistent with the assigned curriculum. Additionally, the researchers used information gathered through implementation research for further analyses, finding that curriculum training, math instructional time, coverage in many math content areas, and instructional approaches were often (though not always) associated with differences in achievement. A subsequent exploratory study examined the associations between instructional practices and student math achievement, generating hypotheses for further research (Clements et al., 2013).

Implementation research can indicate which intervention components and features may be more effective than others, using information from multiple studies. Implementation research provides essential information about the specific components and features of an intervention (or many similar interventions) and their implementation for meta-analyses that combine information across multiple studies. Weiss et al. (2022), for example, examined associations between community college intervention components and impacts on student outcomes. Another meta-analysis examined interventions that address externalizing behavior problems in youth, analyzing 72 studies on family relations and parenting skills, 91 studies on relational approaches, 121 studies on skill-building approaches, 27 studies on behavior management approaches, and 75 studies on academic and education interventions (Wilson et al., 2020). For each of these approaches, the study examined associations between impacts and intervention components gathered through implementation research in each of the studies: intervention content, intervention structure (such as setting, format, standardization, dosage, personnel), and implementation strategies and problems (such as training, supervision, reported implementation problems). Among skill-building approaches, for example, the meta-analysis found associations between improved effectiveness and components of using a staff specialist, using lesson plans, and using content that emphasized conflict resolution skills. For all approaches, the study found positive associations between impacts and three general implementation features: implementation quality, delivery complexity, and provider training or supervision. Results from meta-analyses like these and from informal reviews can then be used in systematic

---

[8] The four curricula were Investigations in Number, Data, and Space; Math Expressions; Saxon Math; and Scott Foresman-Addison Wesley Mathematics.

efforts to test hypotheses and build the evidence base about intervention components and implementation processes.

How this guide is organized

The following sections of the guide discuss key ideas and steps for conducting and reporting implementation research in the context of impact studies:

- **Section I: Establish research questions and begin developing a plan for collecting, measuring, and analyzing implementation data**. This section identifies implementation research questions that apply to most impact studies and discusses early development of plans for collecting, measuring, and analyzing implementation data.

- **Section II: Specify details about the intervention and its implementation**. This section describes ways to conceptualize the intervention and its implementation and to measure implementation features.

- **Section III: Specify details about context**. This section outlines key dimensions of the context in which the intervention takes place and ways to measure those dimensions.

- **Section IV: Specify details about the intervention contrast**. This section describes ways to think about and measure the intervention contrast (that is, the *difference* between what the intervention and control groups receive).

- **Section V: Analyze and report details about the intervention and its implementation as part of the impact study**. This section describes approaches for analyzing and reporting information about implementation.

The information in the guide applies to a range of impact studies, from those testing narrowly focused interventions (for example, reminding students by text message about homework assignments instead of sending reminders through an online learning platform) to those testing multicomponent interventions (for example, whole-school reform). The recommendations presented in the guide (Exhibit 2) are most relevant to prospective experimental studies; however, the guide also provides some recommendations for retrospective nonexperimental designs.

The guide provides definitions, descriptions, templates, and references to additional resources. It draws from a range of examples and studies, including a running example of a hypothetical pre-kindergarten intervention model. This example is a simplified, adapted version of the intervention tested in the *Making Pre-K Count* study conducted by MDRC (n.d.).[9]

---

[9] See Maier & Mattera (2015) for the study's implementation measurement and analysis plan.

Exhibit 2. Overview of the recommendations found in this guide

Recommendations	Steps
1. Establish research questions and begin developing a plan for collecting, measuring, and analyzing implementation data	• Specify research questions about intervention implementation, context, contrast, and associations with impacts • Begin developing a data collection and analysis plan for the study's implementation research
2. Specify details about the intervention and its implementation	• Specify the *planned* components of the intervention in your study: Consider four planned dimensions of each component: content, quantity, mode, and quality • Identify data sources for measuring direct components and support components • Specify how your study will measure intervention fidelity
3. Specify details about context	• Identify context features that are likely to moderate implementation and impacts in study sites • Identify data sources and measures for context features
4. Specify details about the intervention contrast	• Prioritize aspects of the intervention contrast for gathering information • Identify information about and data sources for the intervention contrast
5. Analyze and report details about the intervention and its implementation as part of the impact study	• Address each articulated research question using appropriate qualitative and quantitative methods • Report findings in ways that support their use and future research

Section I. Establish research questions and begin developing a plan for collecting, measuring, and analyzing implementation data

Study teams should formulate research questions about impact and implementation in tandem. Impact and implementation are inextricably linked, and consumers of research benefit not only from information about the efficacy of an intervention, but also information about the tested intervention itself, the context in which the study was conducted, the contrast in intervention receipt that the impact study is testing, and evidence that suggests how aspects of implementation may be associated with observed impacts.

This section of the guide begins with a focus on the early steps of study design, emphasizing the establishment of specific implementation research questions that are complementary to a study's questions about impact. Explicitly stating areas to examine in the form of research questions, instead of as broad categories about which to collect information, helps focus study planning and targets resources toward the highest study priorities. Other issues of study design are also included in this section. To further focus study planning and target resources for implementation research, the guide offers recommendations on developing plans for the kinds of data you will collect, how you will measure various constructs, and how you will analyze the data. Sections II-V of the guide provide further detail about the constructs, measures, data sources, and methods for addressing the research questions and for planning your study.

Specify research questions about intervention implementation, context, contrast, and associations with impacts

This section describes four broad research questions with corresponding examples that detail their implementation, context, contrast, and associations with impacts that are central to impact studies of education interventions.

Early in study planning, researchers should specify implementation research questions (specific examples are shown below). They should start with broad questions about the intervention and its implementation (research question 1), the contexts in which it is implemented (research question 2), the contrast between the tested intervention and the services received by the control group (research question 3), and how intervention components, context, and contrast are associated with impacts (research question 4). Reflecting the IES SEER standard and recommendations on equity (IES, 2022b), include research questions that address equity, drawing on sources such as Cerna & Condliffe (2021), Trainor & Bal (2014), and Woodson (2021).

Broad research questions should address both central tendency and variation. Variation may occur across locations (that is, in different classrooms, schools or other settings, districts,

urbanicity, state, region); time (such as day, week, calendar year, school year); or implementer or participant characteristics (such as age, gender, race, ethnicity, socioeconomic background, education level, prior experience). These variations may be planned and expected, or they may occur incidentally as a consequence of where or how an intervention is implemented.

Research question 1: What components of the intervention were implemented? How did implemented components vary?

Examples of detailed questions: How was the planned intervention adapted to reflect the cultural contexts in study sites? To what extent was the intervention actually offered as planned? What components did intervention group members receive? How did planned, offered, and received components vary over time, across study sites, across students or study participants, or across implementers, with a focus on those that are historically underserved? (See Section II for more information.)

Research question 2: In what contexts was the intervention implemented? How did contexts vary?

Examples of detailed questions: What were the ages, race, ethnicity, and family income of intervention group and control group members? What was the funding structure of campuses in the study? What was the unemployment rate in study communities? What were attendance options and quarantine policies during pandemics? What were the conditions of equity, and structural barriers to equity, in the study sites' programs, schools/institutions, and communities? How do these characteristics vary across study sites or over time? (See Section III for more information.)

Research question 3: What was the intervention contrast? How did the contrast vary?

Examples of detailed questions: How did the frequency and length of meetings with career advisors differ between intervention and control group members? How did the number of sessions, the time spent, and the content covered in after-school reading tutoring sessions differ between intervention and control group members? How did these aspects of contrast vary across study sites, over time, or by characteristics of intervention and control group members, with a focus on those that are historically underserved? (See Section IV for more information.)

Research question 4: What aspects of implemented intervention components, context, and contrast are associated with (or moderate, or mediate) impacts?

Examples of detailed questions: How is the number of meetings with career advisors associated with impacts? How is intervention fidelity in after-school reading tutoring associated with

impacts? How is location (that is, urban, suburban, or rural) associated with impacts? How is intervention contrast in participation in reading tutoring associated with impacts? (See Section V for more information.)

After generating and grouping implementation research questions in each of the broad categories above, ensure they are tailored to your study's features, contexts, and goals. This can include tailoring the questions to the specifics of the intervention tested in your study, the evidence base that you are drawing on, and the study design you are developing.

The example below includes research questions from a hypothetical pre-kindergarten intervention study.

Example: Specifying research questions

A pre-kindergarten intervention model (referred to here and throughout the document as "the hypothetical pre-K intervention") is being tested using a randomized controlled trial. The study is being conducted in 60 pre-kindergarten sites (public schools and community-based organizations) in New York City. Half the sites will be randomly assigned to receive the pre-K intervention, and the other half assigned to the control group, continuing with their typical programming (which may change in response to funding or regulatory changes). The pre-K intervention involves both direct components (small-group instruction and computer lessons for students) and support components (in-person training for teachers and coaching throughout the school year). The study plans to examine impacts on student outcomes at kindergarten entry and again in third grade. Exhibit 3 shows sample research questions for the hypothetical pre-K intervention study.

Exhibit 3. Implementation research questions for hypothetical pre-kindergarten study

Broad implementation research questions	Examples of detailed implementation research questions
Research question 1: What components of the intervention were implemented? How did implemented components vary?	• What intervention components of small-group activities and computer activities did children in intervention group classrooms receive? How did they vary across sites and across student characteristics? • What professional development components were offered to teachers in the intervention group? What did these teachers participate in? How did it vary across teacher characteristics? • What was the intervention fidelity in pre-K classrooms? What was the intervention fidelity for professional development supports? How did intervention fidelity vary across teacher and site characteristics?
Research question 2: In what contexts was the intervention implemented? How did contexts vary?	• What pre-kindergarten policies or requirements (from national, state, city, or district levels) were in place during the study period? • What were characteristics of the sites, teachers, coaches, and children in intervention and control groups in the study? • What were the conditions of equity, and structural barriers to equity, in the study sites and communities?
Research question 3: What was the intervention contrast? How did contrast vary?	• What curricula did teachers use in control group classrooms? On key dimensions, how did the curricula and classroom practices differ between intervention group classrooms and control group classrooms? How did they vary across sites? • What professional development opportunities were available to teachers in the control group? What professional development did teachers in the control group receive or participate in? On key dimensions, how did professional development differ between intervention group teachers and control group teachers? How did it vary across sites?
Research question 4: What aspects of implemented intervention components, contexts, and contrast are associated with (or moderate, or mediate) impacts?	• Was intervention dosage associated with impacts on student outcomes? • Was intervention fidelity associated with impacts on student outcomes? • Was intervention contrast associated with impacts on student outcomes?

Begin developing a data collection and analysis plan for the study's implementation research

An analysis plan for implementation research conducted as part of an impact study guides your study teams' analyses. The plan should:

- Restate your implementation research questions
- Communicate the goals and scope of the implementation research to other study team members and to other stakeholders
- Specify research questions that implementation research will address
- Provide details about the constructs, data sources, and measures for addressing research questions (discussed further in Sections II-V)
- Express any hypotheses about contextual or other features that help or hinder implementation, as well as how variation in implementation might contribute to variation in impacts (discussed further in Sections III and V)

Understanding whether and how impacts vary often involves conducting exploratory analyses using information obtained through implementation research. An analysis plan can describe processes and plans for such exploration. Altogether, analysis plans target resources for the study and help ensure that resources go toward collecting data that will actually be analyzed. Appendix A includes an implementation research analysis plan template.

An early version of the implementation research analysis plan can be in the study proposal, but it should be refined as the study proceeds. In this early version, we suggest pre-specifying certain aspects of the implementation research design (such as defining how your study will measure intervention fidelity), and also describing how the study team will adjust the implementation research analysis plan to address emerging learning (for example, planning to document and measure unexpected context conditions that may arise).

The plan for collecting and analyzing implementation data may evolve as the study proceeds and more detail is needed (for example, developing survey instruments or interview protocols and corresponding analysis plans for them), or as the study team learns about unanticipated practices, conditions, or external shocks (such as a global pandemic) in study sites. Depending on the type of changes, review by Institutional Review Board(s), funders, or other entities (such as the federal Office of Management and Budget) may be needed. Retrospective studies may be limited in the information available to collect, or in the ways in which the information evolves during the course of the study.

Sometimes research questions about factors associated with impact variation appear in separate analysis plans for impact analyses; in other cases, the research team produces a separate companion implementation analysis plan. Regardless of how these plans are presented, team members should ensure that the plans align throughout the study.

Information in the rest of the guide can inform your analysis plan for implementation research in an impact study.

Section II. Specify details about the intervention and its implementation

An intervention impact study must fully describe the intervention that was implemented and received by members of the intervention group. This section explores details of research question 1: What components of the intervention were implemented? How did implemented components vary? It first discusses how to specify the components of an intervention and then identifies potential data sources on the components. The section concludes by providing further detail about fidelity definitions, measurement, thresholds, and interpretation, since assessing **intervention fidelity** (sometimes referred to as intervention **integrity**) is a central step in conducting intervention impact studies.

Specify the *planned* components of the intervention in your study

The impact evaluation you are designing aims to study the effects of an education intervention that is hypothesized to improve specific outcomes. This guide refers to the aspects of the intervention that your study team specifies and studies as its **components**. This subsection describes types of dimensions you can use to specify and describe the components of the tested intervention. Clarity about the intervention itself can, in turn, help clarify your research questions, your plans for data collection, and your plans for analysis.

As noted in the introduction and Exhibit 1, the **intervention model** comprises **direct components** and **support components** that are prespecified by your study team, developers, or implementers. For example, direct components of an early writing intervention for pre-K students might include specific demonstration techniques and visual aids, while support components might include conducting training sessions for teachers and obtaining supplies and resources such as writing surfaces, display boards, pencils, and note cards. Although more research is needed, available evidence suggests that support components such as having standardized, quantifiable training components, manualized components, and well-specified technical assistance models are associated with greater fidelity (for example, Cook et al., 2019; Lyon & Bruns, 2019; and Powell et al., 2015).

Your study's planned intervention model, consisting of direct components and support components, lies along a continuum of three study types:

1. **Developing a new intervention.** You may be designing and testing an entirely new intervention or combining direct or support components from existing interventions in a new way.

2. **Modifying an existing intervention.** You may be evaluating an existing intervention that involves planned modifications, or **adaptations** (Box 2), with a goal of fitting in new contexts, increasing retention, improving feasibility, reducing cost, or increasing satisfaction (Wiltsey Stirman et al., 2019). For example, the City University of New York's

(CUNY) Accelerated Study in Associate Programs (ASAP) is a multicomponent program that aims to increase graduation rates among students in postsecondary developmental education (MDRC, 2015). The intervention needed to be adapted when the ASAP model was planned in Ohio: whereas the CUNY ASAP program had provided students with unlimited MetroCards for the New York City transit system as part of the ASAP financial support component, Ohio colleges planned to provide gift cards for gasoline and food, which were more relevant to students in Ohio than mass transit passes (Miller et al., 2020). Scale-up studies are often this type, as previously tested interventions are tested across settings where planned modifications are likely to occur.

3. **Replicating an existing intervention**. You may be replicating and evaluating an existing intervention, with no modifications, to see if it can be implemented with fidelity and has similar effects with different populations or in different settings.

Regardless of where the intervention model that you are studying lies along this continuum, your evaluation will need to clearly specify the planned intervention for at least two reasons. First, your study team will likely need to work with technical assistance (TA) providers (see Box 3) or local implementers to prepare for implementation of the intervention in the study. Second, your study will need to document what was actually offered and how that compared with what was planned–in other words, to document intervention fidelity (discussed later in this section).

Box 3. Technical assistance in impact studies of education intervention models

Different kinds of TA may be provided when education interventions are tested in impact studies.

What is TA?

While there is not a standard definition of TA, it broadly refers to tailored guidance provided to an organization or community to support learning and capacity-building. Specific TA activities may include "coaching, consulting, modeling, facilitation, professional development, site visits, and referral to informational resources… [in formats such as] individualized-group, onsite-virtual, active (high intensity)-passive (low intensity), and peer-to-peer-directed." (Scott et al., 2022, p. 2)

What TA is a component in impact studies of education interventions?

External providers, the study team, or developers may provide TA to implementers or study participants. When TA is prespecified as part of the intervention model, it may be a direct or support component, depending on the intervention.

Why should implementation research document the use of TA in impact studies?

When TA is part of the intervention model, researchers should document it as a component that is planned, offered, and received, just as they would for any other direct or support component. When TA is offered and received (but not prespecified as part of the intervention model), it should still be documented for context (see Section III). Documenting the presence of TA in a study as part of the intervention model or as a context feature is essential information for future implementers and evidence-builders.

In addition to your study team's own development of and understanding of the intervention, your team can learn about and specify planned components in the following ways:

- **Reviewing the intervention's logic model.**[10] A logic model shows how the components of the intervention relate to each other, to intended outputs, and to intended outcomes. It may also show key context features, and indicate hypothesized moderators and mediators. An intervention's logic model, ideally grounded in theory, helps ensure that the study team, developers, and implementers have a shared understanding of what the intervention involves and what the intended outcomes are. In addition to providing a road map for study planning, data collection, and measurement, it provides a concise depiction of the intervention for consumers of study findings. Box 4 includes additional resources about logic models.

  In logic models, "activities" often refer to direct components, and "inputs" or "strategies" often refer to support components. Because the availability and level of detail of logic models may vary, look beyond the logic model to other sources and resources to ensure that you capture all components. Your study team may need to work with program developers and implementers to construct a logic model if the intervention is new, if a

---

[10] This guide uses the terms "logic model," "theory of change," and "conceptual framework" interchangeably.

logic model does not already exist, or if the intervention was designed to serve a different population.

- **Consulting other documentation about the intervention or its components.** Relevant sources may include previous studies of the intervention or of similar interventions, promotional materials, training manuals, and internal reports.

- **Talking with others knowledgeable about the intervention or its specific components that have been previously studied,** including intervention developers, implementers, and potential participants.

Consider targeted queries to elicit information about plans, assumptions, and previous implementation experiences that will be relevant in the context of your study.

Existing logic models, intervention descriptions, or experiences of developers may reflect ideal or previous study settings. Ensure that the planned components the study team documents are ones actually planned for your study. If your study design involves multiple sites or settings, ascertain and document whether the planned components are

> **Box 4. Resources on logic models**
>
> *Logic Models for Program Design, Implementation, and Evaluation: Workshop Toolkit* (Shakman & Rodriguez, 2015). This toolkit provides templates, exercises, examples, and resource lists for developing and using logic models. https://ies.ed.gov/ncee/rel/regions/northeast/pdf/rel_2015057.pdf.
>
> *Evaluation Plan Template* (Price et al., 2016). Section 3.9 (p. 9) describes how to use logic models in evaluation planning and includes references to resources for developing logic models. https://ies.ed.gov/ncee/projects/pdf/EvaluationPlanTemplate.pdf.
>
> *Guidance for Planning and Reporting your Education Innovation and Research (EIR) Fidelity of Implementation Study* (Goodson et al., 2019). Section 2 (pp. 2-4) describes how to develop logic models and provides a logic model template. Available by request from Abt Associates.

the same or different across study sites. After data collection and analysis, your team will be better positioned to determine whether the differences are substantively important.

Consider four planned dimensions of each component: Content, quantity, mode, and quality

For each component, consider systematically the plans for its **content**, **quantity**, **mode**, and **quality** (Weiss et al., 2014).

As noted above, your study may include one or more planned modifications, or adaptations, to an existing intervention. Model developers may discourage flexibility in implementation of some components and their content, quantity, mode, or quality. For example, some interventions, such as Success for All, have very detailed plans for intervention support components, whereas other interventions, such as YouthBuild, may only partially specify the support or direct components because the model explicitly promotes tailoring to local

contexts.[11] Regardless, documenting adaptation from prior intervention models is a critical step in detailing implementation.

Systematically walk through each dimension described below for each component in your study to define the planned intervention. When a component or dimension is not specified, the study team, developers, or implementers intentionally may be allowing flexibility, may be assuming something unstated about the content, quantity, quality, or mode of the intervention (for example, that the intervention will be implemented by teachers in a classroom setting), or may have overlooked some dimension of the intervention that they want to specify. For example, the time of day when advising takes place may be intentionally unspecified so that advisors and students can arrange times that work for their schedules. When you specify analysis plans, data sources, and measures, plan to collect information about both the specified dimensions and flexible dimensions of the intervention.

Following are descriptions of each dimension, along with examples.

Content: *What are the basic ingredients of each component?*

Examples: The CUNY ASAP intervention includes requirements for enrolling (full-time enrollment and specific timing for developmental courses); specific courses and timing for courses (grouped courses, a student success seminar, and early registration); mandatory student services (advising, career and employment support, and tutoring); and financial support (tuition waiver, free use of textbooks, and a transit card) (MDRC, 2015). A reading tutoring intervention for third graders may involve tutoring sessions and a toolkit with instructional materials featuring targeted reading content for use during the sessions. An early writing intervention for pre-K students may involve different scaffolding techniques such as think-a-louds or visual aids.

Quantity: *How frequently, for how much time, and for what duration will each component occur? What is the planned reach of the component?*

You can use the following categories to specify details about quantity:

- **Frequency:** *How often will the intervention be provided during a given period (such as every day or once a week)?*
  Examples: A developmental education advising component will be offered at the beginning and halfway through each term; a reading tutoring intervention will schedule tutoring sessions twice per week; and an early writing intervention will be scheduled every day that the pre-K class meets.

---

[11] See Manno & Treskon (in press) and Wiegand et al. (2015).

- **Time:** *How much time will a typical encounter take (for example, how many minutes)?*
  Examples: The developmental education advising component sessions will last for 45 minutes; each tutoring session for third graders is scheduled for one hour; and each early writing segment is scheduled to last 30 minutes.
- **Duration:** *Over what time period will the component occur (for example, one semester, one school year, or summer break)?*
  Examples: The developmental education advising component will last up to two years; the planned duration for the reading tutoring intervention and the early writing intervention is one school year.
- **Reach:** *What is the target percentage of intervention group members to receive each component of the intervention?*[12]
  Examples: The developmental education advising component aims to interact at least once with at least 80 percent of intervention group members; reading tutors aim to meet with 95 percent of third-grade students in the intervention group; and the planned reach for the early writing intervention is all students in the pre-K program.

Mode: *How, by whom, and when will the intervention components be delivered (for example, in person or virtual; synchronous or asynchronous; one-on-one or small groups; by a teacher, volunteer, peer, or counselor; before school, after school, or during school)?*

Examples: The advising component of a developmental education intervention will be delivered one-on-one either in the advisor's campus office or virtually via videoconference, on a weekday morning before classes start. The third grade reading tutoring intervention will be delivered virtually, one-on-one, during one of the students' intervention block hours. The intervention does not specify the weeks for starting and ending the intervention during the school year, preferring to leave this to the discretion of local schools.

Quality: *How well will the intervention components be delivered?*

There is not broad agreement on what constitutes "quality" in general or in specific settings. For example, Dane and Schneider (1998, p. 45) distinguish between "quality of delivery"

---

[12] Reach is sometimes referred to as "penetration." Its fidelity can be assessed. For example, if the developmental education advising component reached just 10 percent of those for whom it was intended, the intervention would not be implemented with fidelity on this dimension of the component. The current guide follows Weiss et al. (2014) in including "reach" as an aspect of an intervention's "quantity." Other frameworks, such as Proctor et al. (2011), classify "penetration" as one of eight "implementation outcomes" that also include acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, and sustainability. For a few different reasons, this guide does not address these constructs in detail: First, some of these outcomes, such as cost, are addressed by other SEER standards and resources. Second, information from impact evaluations of education interventions may provide insight into Proctor and colleagues' "implementation outcomes," but they typically are not uniquely designed to do so.

(such as "implementor enthusiasm," "leader preparedness," and "leader attitudes toward the [intervention]") and "participant responsiveness" (such as enthusiasm or engagement of participants). Because the literature on articulating and measuring quality is varied and evolving, this guide follows Weiss et al. (2014) in using a broad definition of "quality."

As Weiss et al. (2014) stated, "Treatment quality is perhaps the most elusive dimension of a treatment package.... The basic idea is that quality services create effective interactions between clients and service providers, promote a high level of client engagement and responsiveness, accurately convey the information they are supposed to convey, stimulate deep personal reflection by clients (especially for therapeutic interventions), get delivered on a timely and predictable basis, and so on" (p. 788).

Examples: Quality might include student engagement and participation, levels of interest and enthusiasm, clarity of presentation of materials, intervention pacing, and instructor knowledge, capabilities, and rapport.

The elusiveness of quality arises from the challenge of articulating exactly what it means for a particular component, and challenges of measuring it even if one can articulate it. Researchers often measure quality using proxies for content, quantity, or mode such as the amount of time spent on a task or in interactions with students. Of course, not all measures of content, quantity, or mode are appropriate proxies for quality; when you specify measures of these other dimensions as proxies, be sure to state specifically in your analysis plans (and later reporting) your justification for doing so.

The tools available to measure quality depend on the nature of the intervention, and they are often focused on particular components. Observation tools that measure the quality of classroom interactions, for example, include the Classroom Assessment Scoring System or CLASS (University of Virginia, School of Education and Human Development, n.d.) and the Early Childhood Environmental Rating Scale (Frank Porter Graham Child Development Institute, n.d.), although research about measuring quality in these settings continues (Maier et al., 2020). Survey tools may provide validated measures of quality that are relevant for some interventions, such as the Community College Survey of Student Engagement (CCSSE), which measures constructs such as student-faculty interaction and active and collaborative learning (Marti, 2008; McClenney et al., n.d.).

Identify data sources for measuring direct components and support components

After specifying the planned intervention model by detailing each component and its content, quantity, mode, and quality, identify data sources that your study can use to collect information about the extent to which the intervention model was implemented as planned.

For the sake of efficiency, including minimizing the burden imposed on others, consider what implementation data are already available. Identifying relevant types of data that you may

need to collect for the intervention you are studying may require web and literature searches, inquiries with other researchers, and discussions with staff at the university system, state, district, or school levels. Through these inquiries, you can learn about the data sources that researchers studying similar interventions have used and may be available in your study sites, and about the kinds of data that are available. Existing data sources most likely to include information about direct or support components of interventions are intervention databases, management information systems (MIS), and curricular materials.[13]

Although existing data may be useful, they are unlikely to be sufficient for information about how the direct and support components of the intervention are implemented. In addition to modifying existing program data or MIS to collect information for the study, examples of data the study team might collect include:

- Logs
- Checklists
- Observational data
- Interviews of administrators, staff, participants, or other stakeholders
- Surveys
- Focus groups

Consider what data to collect and when to collect them, how frequently to collect them, and for what samples. This will depend, in turn, on the implementation schedule and how the study will use the data to address specific research questions. For example, implementation data may be needed over a long period if multiple cohorts participate in the study. However, study teams conducting retrospective studies may be limited in the data they can collect about implementation of an intervention's direct and support components.

It may be helpful for study planning and later reporting purposes to keep track of details about data sources by using different templates for different purposes, as shown in the example below. Different templates focus on describing the direct and support components, describing each data source and what dimensions of intervention components it will collect, and summarizing planned collection across data sources.

---

[13] In some cases, access to the data may require developing data sharing agreements. Districts and schools may have different data sharing requirements, and your team will need to collaborate with sites to determine the requirements, address any concerns about data collection, and communicate procedures to protect personally identifiable or sensitive data (see Neild et al., 2022, for more information).

Example: Specifying dimensions of direct and support components in the study and plans for collecting information

Study teams can use an analysis plan to specify the key direct and support components and methods to measure them. Exhibits 4 and 5 provide templates, and example entries for the hypothetical pre-K intervention, that study teams can adapt to include in their own analysis plans.

- Exhibit 4 is a template for listing and describing the dimensions of each direct component and support component specified as part of the intervention model, and the data sources that the team will use for each.

- Exhibit 5 is a template for describing the data collection schedule for each data source in the implementation study.

Exhibit 4. Template for describing and collecting information for each planned intervention component

Component	Direct (D) or Support (S)	Dimension				Data sources for intervention group				
		Content	Quantity	Mode	Quality	Docu-ments	Logs: Teacher	Obser-vations	Logs: Coach	Inter-view: Teacher
Small-group student sessions in classroom	D	Modules of math concepts and strategies, with progression over modules	Weekly module: 5-minute introduction and modeling with 20-minute student activities	In-person groups of 3-4 students	Teacher-student interaction	✓	✓	✓		✓
List any additional direct components and their dimensions.										
Coaching sessions for teachers	S	Coaching session to review previous month's activities, preview next steps, answer questions, provide technical assistance	Monthly 1-hour meeting	In-person or virtual, one-on-one meetings; follow-up by email as needed	Coach-teacher interaction	✓	✓	✓	✓	✓
List any additional support components and their dimensions.										

Exhibit 5. Template for listing data sources and data collection schedule

Data source	Respondent	Provides information about...	Fall 2023	Spring 2024
Documents: Training materials	Training provider	Planned training content	August	
Observations: Training	Independent observer	Training sessions	August	
Logs: Coach	Coach	Coaching sessions with teachers	Weekly	Weekly
Documents: Curricula	Content provider	Planned curriculum	August	Ongoing
Logs: Teaching/classroom	Teacher	Teaching and classroom activities	Weekly	Weekly
Observations: Teaching/classroom	Independent observer	Teaching and classroom activities	Sept/Oct	April/May
Interview: Teacher	Teacher	Teacher, classroom activities, and coaching experiences		April/May

Specify how your study will measure intervention fidelity

Evaluations conceptualize and measure fidelity in various ways (Dhillon et al., 2015; Hill & Erickson, 2019; Nelson et al., 2012). Given these numerous conceptualizations, it is important to ensure that all members of your study team and other stakeholders and audiences have the same understanding of the term. As noted in the introduction, the guide defines **intervention fidelity** as the correspondence between the *planned* intervention components and the *offered* intervention components.[14] Higher intervention fidelity indicates greater correspondence between dimensions of the planned intervention components and dimensions of the implemented intervention components.

Boulay et al. (2018), who provided TA and summarized findings for 67 Investing in Innovation (i3) Fund evaluations, observed, "Though there is a growing consensus in the field that evaluators should develop logic models and measure fidelity of implementation, there is disagreement about what constitutes valid measurement of and thresholds for adequate fidelity" (p. 15). The authors recommend that any fidelity measure should at least include the following:

---

[14] As noted in an earlier footnote, fidelity is one of eight "implementation outcomes" specified in the implementation science literature (Proctor et al., 2011).

- A set of measurable indicators of implementation fidelity for each planned direct and support component
- A way to combine those indicators into an overall score for each component
- A threshold that defines adequate fidelity[15]

It may be possible to assess intervention fidelity for each dimension (content, quantity, mode, and quality) of every direct and support component of the intervention model. Often though, it is difficult to ensure robust measurement of every dimension for all intervention components given other study constraints. In these instances, study teams should prioritize measuring the dimensions of components that are central to the intervention's theory of change. Describe how you will assess measures for validity, such as using external reviewers or calculating correlations to ensure that associated measured constructs that should be associated actually are; and how you will assess measures for reliability, such as through Cronbach's alpha for internal consistency of items in a scale, or Cohen's Kappa or other measures for inter-coder reliability.

For example, TA provided to IES' Education and Innovation Research (EIR) grant recipients for evaluation purposes specifies that intervention fidelity should be measured for the entire implementation sample and assessed at the level of analysis where the component is implemented (for example, teacher, school) (Goodson et al., 2019). Box 5 shows an excerpt from its general guidance for measuring intervention fidelity, and Appendix B shows an excerpt from its template for assessing intervention fidelity. Sources such as Sanetti & Collier-Meek (2019) also provide guidance on measuring intervention fidelity; and Nelson et al. (2012) provide more in-depth guidance, including how to assess reliability and validity, and how to combine indices.

---

[15] Ideally, prior evidence informs the threshold for intervention fidelity. If strong evidence is not available, the team should prespecify thresholds based on the developer or researchers' experiences and beliefs regarding implementation of the intervention. Sources such as Goodson et al. (2019) provide further guidance about establishing fidelity thresholds.

Box 5. Excerpts from guidance for EIR grantees for measuring intervention fidelity

Guidance for Planning and Reporting Your Education Innovation and Research (EIR) Fidelity of Implementation Study, Section 3 (pp. 1-14) provides guidance on measuring intervention fidelity and provides an example and template. The following excerpt from the *Guidance* describes part of what grantees must do to measure fidelity separately for each component of the intervention:

1. The **unit** at which the key component is delivered and measured–teacher, grade, school, district, or program. The unit of delivery and measurement of a key component should be consistent with the specification of the measurable indicators for the component…. For example, if the key component is *coaching for teachers on the instructional approach,* the individual teacher is the unit of delivery and measurement.

2. The set of **measurable indicator(s)** that define the key component. For example, if the key component is *coaching for teachers on the instructional approach,* the indicators should specify how the component will be measured at the individual teacher level, including features such as the expected number of hours of coaching per teacher, the number of post-coaching feedback sessions a teacher is expected to receive, etc.

3. The **approach to scoring** each indicator. The fidelity measure should specify how each indicator will be scored. For example, specify whether the indicator will be scored as a continuous number (for example, "number of hours of coaching"), as a binary number ([for example], "0 = teacher received less than 15 hours of coaching, 1 = teacher received at least 15 hours of coaching").

Source: Excerpted from Goodson et al. (2019, pp. 3.1-3.2).

You may decide how to measure intervention fidelity in your study after considering how others have conceived of and measured it. Box 6 includes links to just a few of many possible examples. An edited volume (Meyers & Brandt, 2014) focuses on fidelity measurement in i3 evaluations, including how evaluators addressed challenges such as identifying thresholds.

Your study may assess fidelity using measures already specified by developers, replicating measures from the same or similar interventions in prior studies, or developing new fidelity measures (Meyers & Brandt, 2014). For example:

Box 6. Examples of fidelity measurement

Investing in Innovation (i3) Fund evaluation (Boulay et al., 2018, pp. 17-18): https://ies.ed.gov/ncee/pubs/20184013/pdf/20184013.pdf.

Reading Recovery evaluation (May et al., 2016, pp. 70-71): https://repository.upenn.edu/cgi/viewcontent.cgi?article=1089&context=cpre_researchreports.

- **Use measures already specified by developers or previous studies.** This approach is likely to be most applicable among studies that are replicating an existing intervention. For example, Bradshaw et al. (2012) assessed fidelity of schoolwide positive behavioral interventions and supports (PBIS) using the Schoolwide Evaluation Tool, a fidelity measure developed by Horner et al. (2004), who also developed the PBIS framework.

- **Modify measures from previous studies**. This approach will be most applicable among studies that are modifying an existing intervention. For example, the Making Pre-K Count study team modified the Building Blocks classroom observation fidelity measure to account for an observation conducted in one morning instead of multiple visits over two months (Maier & Mattera, 2015).

- **Develop new measures.** This approach will be most applicable among studies that are developing a new intervention or measuring fidelity of an existing intervention for the first time. For example, the Diplomas Now study team developed two metrics: a continuous score that examined all nine components (such as curriculum for college readiness, student case management) and many subcomponents; and a categorical rating that focused on a subset of components that Diplomas Now staff members at the national level hypothesized to be most critical for effectiveness (Corrin et al., 2014).[16]

If your study team is modifying existing measures or developing new measures, make sure to follow standard measurement practices such as tapping into multiple dimensions, checking for reliability and validity, and documenting measurement conditions.[17] If your study is examining associations between intervention fidelity and impacts in a multisite study, it will be important to measure intervention fidelity separately in the sites where separate impacts will be estimated. If feasible, fidelity scores can provide more information than binary indicators about the presence or absence of component dimensions.

---

[16] Corrin et al. (2014) referred to the nine broad components as "model inputs." The study's Appendix Table B.2 shows subcomponents for each input and how the study defined fidelity for each.

[17] Sources such as Nelson et al. (2012), Hulleman et al. (2013), Harn et al. (2013), McCormick & Maier (2018), and Sanetti & Collier-Meek (2019) discuss these practices.

Section III. Specify details about context

An intervention impact study provides rigorous evidence about whether an intervention is effective for a particular population in particular organizational and environmental settings. This section explores details of research question 2: In what contexts was the intervention implemented? How did contexts vary? It discusses two steps: (1) identifying context features in study sites that are likely to moderate implementation and impacts; and (2) identifying data sources and measures for context features.

Context features in a study include characteristics of the implementing organization and systems, study participants, and the external environment in which the intervention is implemented (see Exhibit 1). The Consolidated Framework for Implementation Research (CFIR) (Damschroder et al., 2009), the Exploration, Preparation, Implementation, and Sustainment (EPIS) framework (Aarons et al., 2011; Moullin et al., 2019), and Domitrovich et al. (2008) identify potentially relevant context features. The context features described in this section and the support components described in Section II are similar. Context features are moderators that are not part of the intervention model, whereas the support components are specified as part of the intervention model.[18]

Context is relevant for studies at any stage. This includes studies that focus on developing a new intervention or modifying an existing one. Attention to context is particularly salient when studying programs adapted to fit new contexts, such as serving participants with different social identities and lived experiences.

Identify context features that are likely to moderate implementation and impacts in study sites

This subsection describes a number of constructs that may be relevant context features in your study. In addition to drawing on these lists for ideas, consult impact studies of similar intervention models, other study team members, intervention developers, and other stakeholders to help identify contextual features that are relevant for the intervention you are studying.

It is important to think about characteristics that are most likely to moderate implementation or impacts of the intervention in study sites rather than taking a "kitchen sink" approach to collecting every possible measure available. All things being equal, it is most useful to focus on characteristics that (1) are likely to shed light on contexts that make impacts more or less likely to be attained; (2) can be addressed through enhanced supports for implementation; or (3) point to opportunities to improve the intervention model itself. For example, teachers' years of experience or burnout levels may affect their willingness or ability to participate in an intervention that involves working with a mentor and implementing new practices or

---

[18] The corresponding terms in the implementation science literature are "determinants" (referring to context features) and "strategies" (referring to support components).

curricula in the classroom. This may suggest settings in which an intervention is not appropriate, or might need to be paired with complementary teacher supports to make implementation feasible.

Some of the guiding questions about equity suggested by Cerna & Condliffe (2021) are particularly relevant to issues of context. The authors ask:

- "What are the baseline conditions of equity, and of structural barriers to equity, in the ecosystem where the project will take place? (p. 3)

- What key sociocultural characteristics surrounding the project may play a role in program implementation? (p. 4)

- Does the implementation research seek multiple perspectives from stakeholders and communities about existing inequities or biases that they may feel are being perpetuated?" (p. 4)

Example: Specify context features and plans for collecting information about them

After your study team selects study sites, you can begin recording context in a format such as that in Exhibit 6. It shows context features for the hypothetical pre-K intervention study conducted in New York City.

Exhibit 6. Template for documenting context features of the intervention

Context feature	Unit*	Description	Data source
Policy context	6, 7	State and city policies related to pre-kindergarten curricula	Document review
Location	6	NYC borough where study site is located	Document review
Organization type	3	Public school, community-based organization, or Head Start	Document review
Organization characteristics	3	Student characteristics; teacher turnover rate	Document review
Other professional development	3	Training or coaching available that is not part of the intervention model	Interview: Teacher
Readiness for change - teacher	1	Readiness for changing teaching practices	Survey: Teacher
Background characteristics - teacher	1	Demographics, education levels, work experiences	Survey: Teacher
Background characteristics - coach	8	Demographics, education levels, work experiences	Survey: Coach

* Unit values: 1 = Teacher; 2 = Classroom; 3 = School/organization; 4 = District; 5 = System; 6 = Community; 7 = State; 8 = Other.

A number of context features may be relevant for your study (see Box 7 for the specific features that the What Works Clearinghouse emphasizes), including:

- **Characteristics of the broader external environment.** These are macro-level or outer-setting characteristics. They include policies (such as whether the state has required schools to develop standards for social-emotional learning); financing structures (many districts face challenges financing their programs); characteristics such as the education levels of the local labor force and labor availability; historical or structural factors, such as segregation; and unanticipated occurrences such as the COVID-19 pandemic or school shootings.

- **Characteristics of the district, school system, organization, classroom, or inner setting.** These include the following factors (Damschroder et al., 2009;

Domitrovich et al., 2008; Society for Implementation Research Collaboration, 2021):

- Alignment between the intervention and the school's mission
- Decision structure (that is, centralized, dispersed)
- Resources (for example, knowledge, equipment, staff time, space)
- Expertise of personnel
- Administrative leadership
- Organizational or school culture (Powell et al., 2021)
- Organizational or school climate (Institute of Behavioral Research, 2009; National Center on Safe Supportive Learning Environments, 2022; Powell et al., 2021)[19]
- Readiness for change (in psychological terms and structural terms) (Weiner, 2009; Weiner et al., 2008)

Box 7. Context features listed in the What Works Clearinghouse Reporting Guide for Study Authors

- County or state, if in the United States
- Urban, rural, suburban, or town setting
- Number of schools, postsecondary institutions, or education sites
- School type–charter, parochial, public or private–and format–in person, online, before or after school
- Postsecondary institution type–two-year, four-year, private
- Other educational site–center, home-based
- Classroom type, including general or inclusion, self-contained special education, or designated English language development
- Other school, institution, or site characteristics, including enrollment, Title I status, magnet, aggregate student characteristics
- Educator characteristics
- Student-level characteristics

Source: Adapted from IES (2021b).

---

[19] EdInstruments compiles measures of school climate, listing 26 measures as of March 2023 (EdInstruments, 2023).

- Characteristics of the school (such as number of students, aggregate student characteristics such as race, ethnicity, socioeconomic status, absenteeism, mobility)
- Other initiatives being implemented in the district, school, or organization at the same time as the intervention you are studying

- **Characteristics of individuals** (for example, of students, teachers, staff, leaders). These include the following:
  - Demographic characteristics
  - Educational background
  - Professional background and work experiences
  - Attitudes and beliefs
  - Social-emotional characteristics (for example, burnout, depression, anxiety)
  - Psychological factors, such as readiness for change

The broader conditions that place your study in historical and geographic context may be particularly important to consider. For example, a study conducted during the 2020-2021 academic year should note specifically that the COVID-19 pandemic was occurring during this time, and in most states causing major disruptions to school operations that resulted in widespread shifts to virtual learning. Without describing these important factors, future consumers of the study's findings may not interpret your findings appropriately. The importance of documenting context features is underscored by a recent study conducted in a school district near Ferguson, Missouri, where Michael Brown was shot and killed by police in 2014 (Herman et al., 2022). Protests and school closures followed. The randomized controlled study of a classroom management intervention found that the intervention had greater effects for Black teachers than for non-Black teachers before Brown's killing but not after. Effects on student achievement scores for both White and Black students were small in magnitude and statistically significant before Brown's killing; afterward, impacts continued for White students but not Black students. This example shows how historical events may be associated with the findings for a randomized controlled study.

Identify data sources and measures for context features

Many existing secondary data sources contain potentially relevant context information for education studies. Much of the guidance about data sources described in Section II applies for context features, including recording information in a template. Existing data sources with information about context features include the Federal Reserve Economic Data (2023) and sources referenced in Tipton & Olsen (2022), such as the Common Core of Data (IES, 2023a), the education demographic and geographic estimates from the American Community Survey (IES, 2023b), policy or program information from the Education Commission of the States (ECS, 2023) and from state, district, and school websites or MIS.

Example: Context features for a hypothetical pre-kindergarten intervention study

Various context features can affect how coaches provide professional development to teachers, the extent to which teachers engaged in that professional development, and how teachers implemented the curriculum in the classroom. They include:

- **Community-level factors** of borough location and pre-K policy environment in New York City during the study

- **Site-level factors** such as program type and funding streams

- **Classroom-level factors** such as aggregate student characteristics (for example, percentage of dual language learners in the classroom or number of children who have faced disciplinary incidents), quality of classroom management, and extent of role equity in the teaching team

- **Teacher-level factors** such as educational background, prior training and teaching experiences, readiness to change and motivation to implement, attitudes and beliefs toward math, burnout and mental health, quality of the teaching team's relationship

- **Coach-level factors** such as coach attitudes and beliefs toward relationship quality between math teacher and coach

- **Child-level factors** such as English language learner status, gender, and baseline test scores

You may need to collect primary data for some measures of context features such as school climate, building or structural supports, or individual or organizational readiness for change. Use available validated and reliable measures, if possible. Researchers should also prioritize measures that incorporate inclusive language (American Psychological Association, 2021). In addition to modifying existing program data or MISs to collect more information for the study, you can collect these and other context features in the following ways:

- Surveys of students, staff, or other stakeholders

- Activity logs

- Checklists

- Observational data

- Interviews of administrators, staff, participants, or other stakeholders

- Focus groups

Study teams conducting retrospective studies–whether lottery-based experiments or nonexperimental studies–may be limited in the data they can collect about context features, especially those at the classroom or individual levels.

Section IV. Specify details about the intervention contrast

An intervention impact study provides rigorous evidence about whether an intervention is effective relative to the counterfactual. This section explores details of research question 3: What was the intervention contrast? How did the contrast vary? Two steps are relevant for this discussion: (1) prioritizing aspects of the intervention contrast you want to gather information about; and (2) identifying information about and data sources for the intervention contrast.

Intervention contrast can vary widely across studies, locations, time, and even across sites within a study (see, for example, Lemons et al. 2014). Even if an intervention has a strong logic model and is implemented with high fidelity, impacts on student outcomes may not be detected if control group members have access to and use similar strong services or supports. Yet researchers and stakeholders often focus on intervention fidelity and what the intervention group actually receives and devote less attention to the contrast between services received by the intervention and control groups.

Prioritize aspects of the intervention contrast for gathering information

Because the intervention contrast depends on both the study design and contexts, and because it is unlikely you will be able to measure every single aspect of contrast, plan to prioritize. To identify key contrast components and dimensions, consider the following:

- **The logic model.** Which components–and which dimensions of those components–do you hypothesize are central for improving student outcomes for the population you are studying? These components may be of greatest interest when interpreting the presence or absence of impacts. Prioritize the intervention contrast for these central components. In considering which components or their dimensions are central, consider which are most likely to lead to substantively large effects. Also consider the strength of the evidence from prior research that these components or their particular dimensions are likely to drive impacts.

- **The counterfactual environments or intervention contrasts documented in previous studies.** If your study is examining a modification or replication of an existing intervention, check previous studies to see which aspects of contrast were measured and their relevancy for the sites in your study.

- **The availability of services or programs in study sites that are similar to the intervention you are studying**. At one extreme, nothing close to the intervention you are studying is or will be available to control group members throughout the study. At the other extreme, close substitutes are available currently and more are due to become available during the study. Of course, many possible realities exist between these two extremes, and it is not possible to fully anticipate all changes in counterfactual conditions. Prioritize measuring the intervention contrast for the central components, especially

where counterfactual offerings are similar to the intervention. Ensure that all study team members understand the critical role of intervention contrast and how it contributes to measured intervention impacts (see Exhibit 1 presented earlier in this guide). Also ensure that team members recognize counterfactual conditions that could affect contrast, such as policies or program offerings, curricula, or other initiatives.

- **The difficulty or cost of reliably measuring intervention contrast.** What data about the intervention contrast are likely to be available for both the intervention and control groups? What do you know about data quality and cost of data acquisition? All else equal, prioritize intervention contrast measures for which data from the same source are available for both intervention and control groups, and for which data quality is high and data acquisition costs are low. However, for priority components or dimensions (identified above), devoting additional study resources to obtaining high-quality information about the intervention contrast may be necessary.

Example: How changing policy context can affect intervention contrast

State or local policy changes affecting pre-kindergarten curricula are likely to affect the intervention contrast for a study such as the hypothetical pre-K intervention. For example, the District of Columbia revised its early learning standards in 2019 to align with the Common Core (Office of the State Superintendent of Education for the District of Columbia, 2023). If the changes induced by such policies result in more similarity to a tested intervention, then the intervention contrast will narrow. The Making Pre-K Count study encountered such a situation, discussed further in Mattera et al. (2018).

Identify information about and data sources for the intervention contrast

Researchers often focus their time and resources on measuring aspects of the intervention being tested and do not consider the extent of the data they may be able to access from services received by the control group. Intervention components provide a lens through which you can view alternative services and supports. For the direct components and support components of the intervention with the highest priority, plan to collect information to determine if the same or similar services or supports are available to control group members. Take the following steps to look systematically for potential counterfactual services and supports:

- **Consult prior evidence.** In previous studies, what kinds of alternatives have control group members had access to? How are they similar to or different from those available in the current study?

- **Talk with leaders, staff, and potential participants at study sites or other contacts in the local area.** Learn about the service offerings and supports that are available in counterfactual sites. For example, if the intervention is for an after-school peer tutoring program, learn whether there are other in-school or after-school tutoring programs available in study sites, and whether tutors are peers or staffed in other ways.

- **Check publicly available information from schools, districts, and, when relevant, developers of the study intervention or similar ones.** The leaders or staff you talk with may forget to mention a relevant alternative service or not mention it because they do not think it is relevant. Check school, district, or developer websites for similar interventions, the latter of which sometimes list districts or schools that have adopted the developer's intervention. For example, if you are conducting an evaluation of a developer's after-school peer tutoring intervention, the developer may know about other tutoring options available in the study sites.

When identifying data sources that you will use to collect information on control group members, assess what is both possible and reasonable to prioritize. Consider the following:

- **Using the same data sources and measures for both intervention and control group members whenever possible**. This will ensure that any differences between the groups can be attributed to actual differences in services and not to the use of different data sources or measures. For example, suppose the study obtains information about intervention components through daily logs that teachers in the intervention group complete but obtains information about counterfactual services or programs through a single survey that teachers in the control group complete at the end of an academic year. In this scenario, recall error is likely to be greater in the survey than in the logs, and the information will not be comparable. You may have information from more data sources about the intervention than about the control group, but to the extent possible ensure that there are common data sources across the two groups.

- **Collecting valid and reliable information throughout the project on these services**. These data should be high-quality, valid, and reliable to the extent feasible. For example, classroom observation data should be collected by trained coders using validated protocols with known psychometric properties.

  It is critical to monitor counterfactual services and offerings throughout the study (Hamilton & Scrivener, 2018; Weiss et al., 2014; Hulleman & Cordray, 2009). There are at least two reasons to do so. First, the counterfactual environment may change (such as new policies, programs, curricula, leaders, staff); hence, the intervention contrast may not be stable over time. Although the study's ability to generate a causal estimate remains intact, the potential magnitude of the effect may change. Second, without early and continued focus on the counterfactual and contrast, your study team may realize only when it is too late that the study should have collected information on services that both the intervention and control groups received.

- **Using measures that are not highly specific to the intervention.** Instruments or protocols that the study team uses to assess fidelity to the intervention may not be appropriate for measuring services or supports control group members receive because the instruments focus too narrowly on the intervention itself. As Nelson et al. (2012) stated, "Fidelity measures must contain at least some indices that are sufficiently general ([that is], construct-based) to be applied to the control condition as well as the treatment condition" (p. 377). For example, suppose an intervention focuses on teachers' use of guided instruction. An observation protocol designed specifically to identify guided instruction might not reflect exposure of control group members to didactic teaching instruction. That narrow focus is appropriate for using such an instrument to assess whether the intervention was implemented as intended. But if used to assess counterfactual services, the instrument may not capture practices that are substantively similar to the intervention practices (in this case, exposure to active teaching of any kind). Using the fidelity measures to examine the intervention contrast could understate the services the control group received, thus overstating the intervention contrast.

Example: Specifying and documenting intervention contrast for direct and support components

Study teams can use an analysis plan to specify whether and how they will collect information about components for the control group that are most similar to the intervention's direct and support components, and to articulate methods to measure these components. Exhibit 7 provides a template and example entries for the hypothetical pre-K intervention that study teams can adapt to include in their own analysis plans.

Exhibit 7. Template for tracking data sources to measure intervention contrast

Component	Description	Documents		Logs: Teacher		Observations		Logs: Coach	
		Interv.	Control	Interv.	Control	Interv.	Control	Interv.	Control
Small-group student sessions in classroom	Modules of math concepts and strategies, with progression over modules; weekly 5-minute introduction and modeling with 20-minute student activities; in-person groups of 3–4 students	✓	✓	✓	✓	✓	✓		
List additional direct components and their dimensions.									
Coaching sessions for teachers	Coaching session to review previous month's activities, preview next steps, answer questions, provide technical assistance; monthly 1-hour meeting; in-person or virtual, one-on-one meeting with follow-up by email as needed	✓		✓	✓	✓		✓	
List additional support components and their dimensions.									

Section V. Analyze and report details about the intervention and its implementation as part of the impact study

Findings from an intervention impact study are most useful for future practice and research when they are accompanied by information about the intervention's components, contexts, and contrast, and how these factors are associated with impacts. This section explores examples of how to analyze and report information from your study. There is not a standardized way to conduct analyses and report findings related to the four broad implementation research questions found in this guide. The importance of each research question depends on the nature of the intervention, the study contexts, the range of decisions the study is designed to inform, and the audience of a specific study product.

Data sources and general analysis and reporting approaches for addressing implementation research questions

Research questions 1, 2, and 3 pertain to implementation, context, and contrast and can be addressed using qualitative and quantitative data sources and analytic approaches. The primary data sources for these three research questions are identified in Sections II, III, and IV. Measures and analyses used to inform research questions 1-3 can then be used to inform research question 4 (associations with impacts), which is discussed in another subsection below.

Conducting analyses of qualitative data. Whether a study team is analyzing open-ended survey items, data from activity logs, or transcribed or recorded interview data, qualitative analyses may take a **deductive** approach (in which the study team identifies topics or themes before data collection and coding), an **inductive** approach (in which the team identifies themes by analyzing the data), or a combination of both approaches.[20] Exhibit 8 shows an excerpt from Gray et al. (2022) that describes how their team coded the collected information, using both deductive and inductive approaches. This approach helped to identify key themes relating to their implementation questions.

---

[20] Detailed guidance and further resources about qualitative methods are available in Miles et al. (2019), Patton (2014), Cohen & Crabtree (2006), and Wagner & Wright (2022), among others.

Exhibit 8. Deductive and inductive approaches used in *Zoology One* study

To understand the reasons for variation in teachers' implementation of *Zoology One,* we analyzed data from interviews with 12 treatment teachers in Cohort 1 and 16 treatment teachers in Cohort 2, representing a total of 21 schools. We developed and used codes that emerged both inductively and deductively from the *Zoology One* logic model or interviews. We applied the codes to randomly selected transcripts, then discussed discrepancies to arrive at common understandings until 80% reliability was reached. To enhance the validity of our findings regarding factors contributing to variation in teachers' implementation of *Zoology One,* the research team applied three analytic strategies to the data coded with the "implementation factors" main code. First we counted how many transcripts included a particular sub-code of the "implementation factors' main code (e.g., "coach") at least one time. This was a way to determine the prevalence of that sub-code within the overall sample. Next, we counted sub-codes within individual transcripts, and tallied the number of transcripts within which a given sub-code appeared most frequently. Last, we looked at which sub-codes emerged most often across all transcripts. After applying all three analytic methods to the data coded "implementation factors," we ranked the sub-codes by frequency across all three methods to derive our key themes.

Source: Recreated from Gray et al. (2022, p. 13).

Conducting analyses of quantitative data. Analyses of quantitative data can provide useful insight into implementation issues, complementing qualitative analyses. Before using quantitative data to generate summary statistics and conduct analyses about implementation, context, and contrast, study teams should conduct data quality checks and employ data reduction strategies such as those described in Appendix C.[21]

Of particular interest for implementation research, analyses might examine how two contextual variables relate to each other in the study sites, such as whether urban sites are more likely than non-urban sites to implement the intervention in a public charter school rather than a traditional public school. Analyses also might examine associations among contextual variables and implementation variables. For example, the study could examine how the type of school or implementing organization is associated with intervention fidelity or how the urbanicity of a site is associated with the intervention contrast.

Conducting analyses using multiple or mixed methods. Ideally, quantitative and qualitative information can be synthesized to provide a holistic picture of implementation. "Multiple methods" refers to analytical approaches in studies that are both qualitative and quantitative. "Mixed methods" refers specifically to study designs and analyses in which one form of data collection and analysis intentionally informs another. For instance, a team may primarily use qualitative data to develop themes, then identify areas of overlap with quantitative data and perform a selective analysis to help fill out its story. Conversely, the team could begin by analyzing quantitative data and let these findings drive the qualitative

---

[21] Detailed guidance about quantitative methods is available in sources such as Remler & Van Ryzin (2021).

analysis. Teams could also use a more integrated approach, coding both at the same time and then aligning the qualitative and quantitative findings to see where they agree and where more analysis is needed to explain a mismatch.[22]

Reporting findings for each research question about intervention components, context, contrast, and associations with impacts. Your impact study–no matter how many sites or sample units–will have a greater chance of building evidence for the field and being informative for potential implementers of the tested intervention if it clearly and succinctly conveys information about intervention components, contexts, contrast, and associations with impacts. The next section discusses each of the broad research questions articulated in Section I and provides suggestions for analyzing and reporting findings.

Analytic and reporting approaches for studying implementation of intervention components and their variation (research question 1)

The first research question helps users of research understand what components were planned, what was actually implemented, how it was implemented, and the degree to which implementation varied. As discussed in Section II, your study can collect data on implementation of components in several ways and along different dimensions (content, quantity, mode, and quality). Similarly, information about implementation components can be analyzed and reported in many ways. This can include summary boxes, logic models, detailed text, tables, or other displays. A concise summary included in an introduction or executive summary helps ensure that the information is easily accessible. Exhibit 9 shows how May et al. (2016) presented high-level takeaways about intervention fidelity in the executive summary of their i3 evaluation of Reading Recovery.

---

[22] Detailed guidance about designing and conducting mixed methods analyses is available in Creswell & Plano Clark (2017), and overviews are available in Creswell (2021), Palinkas et al. (2011), and Wagner & Wright (2022).

Exhibit 9. Example of reporting summary findings about intervention fidelity and implementation

CPRE/CRESP assessed fidelity to the Reading Recovery program model in the schools involved in the scale-up over the four years of the evaluation. Fidelity data were collected annually from Reading Recovery teachers and teacher leaders, and analyzed for consistency with the Standards and Guidelines of Reading Recovery in the United States, 6th Edition (Reading Recovery Council of North America, 2012). Specifically, we assessed fidelity to the program model in four areas: staff background and selection; teacher leader and site capacity; Reading Recovery teacher training and professional development; and one-to-one Reading Recovery lessons. Our analysis revealed strong fidelity to the program model in all of these areas and all years of the scale-up. This suggests that the intervention was delivered as designed to the students in the scale-up, and that teachers delivering Reading Recovery lessons were properly trained. In total, the results of the fidelity analysis support the validity of our impact findings.

Source: Recreated from May et al. (2016, pp. 3-4).

As another example, the eMINTS report on professional development included tables contrasting planned and delivered intervention components (Meyers et al., 2015). Exhibit 10 illustrates how the study compared the components planned versus those implemented for schools assigned to each of the two intervention conditions. Cells show the number of instances of each component listed in the "Criteria" column across all schools in the column. For example, the first row of the table under the "Teachers" heading indicates that the two intervention arms received all 47 of the planned eMINTS professional development sessions, whereas the last row of that section shows that 85 of the 99 planned portfolio reviews occurred. This table provides a starting point in understanding implementation fidelity. The remainder of the report includes more specific details on the degree to which components were implemented as planned, variation in implementation across schools, and challenges schools faced in implementing with fidelity (for example, challenges related to professional development and technology use).

Exhibit 10. Example of reporting findings about intervention components

Program component	Year	Criteria	All treatment schools (N = 38)		eMINTS (N = 18)		eMINTS + Intel (N = 20)	
			Planned	Delivered	Planned	Delivered	Planned	Delivered
Teachers								
Professional development sessions	Years 1 and 2	47 PLC sessions	47	47	47	47	47	47
Coaching visits	Years 1 and 2	20 coaching visits (10 per year)	1,980	1,657	880	762	1,100	895
Online access to curriculum and professional development materials	Years 1 and 2	All teachers receive access	99	99	44	44	55	55
Portfolio reviews	Spring Year 2	All teacher portfolios are reviewed and scored	99	85	44	39	55	46
Principals								
Professional development sessions	Fall Years 1 and 2	Two 2-day sessions (one session per year)	2	2	2	2	2	2
Walk-throughs	Fall and spring Year 1; fall Year 2	Three school walk-throughs with a certified eMINTS specialist (two in Year 1 and one in Year 2)	114	112	54	53	60	59
Technology coordinators								
Webex sessions	Fall and spring Year 1; fall Year 2	Three Webex sessions (two in Year 1 and one in Year 2)	3	3	3	3	3	3

Source: Recreated from Table 3.1 in Meyers et al. (2015).

Documenting implementation need not be a complex endeavor. Tables with descriptive summary statistics can provide critical implementation information about content, quantity, mode, and quality. For example, if you are conducting a formative evaluation, you may present tables that show information such as the frequency of logging into a platform; percentage of professional development modules attended; or average satisfaction ratings. For a study that is exploring the scale-up of an intervention, stakeholders will want to see

comparisons of fidelity across different implementation settings (for example, classrooms or schools within a district).

Summary measures may also be created from multiple items in surveys, logs, or observation tools. For example, Fountain et al. (2016) created summary measures of intervention fidelity in their scale-up study of Citizen Schools' Expanded Learning Time (ELT) model. They created an index linked to core principles of effective implementation of the ELT model that integrated information from multiple data sources into a summary measure describing variation in the schools' implementation efforts. Exhibit 11 provides an excerpt of how they reported this information.

Exhibit 11. Example of implementation index constructs and indicators from evaluation of Citizen Schools' ELT model

Construct Fidelity of implementation reflects...	Indicators
1. Planning	a. Ongoing planning between Citizen Schools and [host] school
2. Leadership	a. Engaged leadership
	b. Support for ELT
3. Data collection	a. Data sharing between 1st and 2nd
	b. Citizen School aligned with data results
4. Training and professional development (PD)	a. Training for Citizen School staff
	b. Increased responsibility for fellows
5. Family and community	a. Families are engaged and informed
6. Alignment and coordination between partner school and Citizen Schools	a. Alignment of content and behavior standards
	b. Nature and frequency of communication
	c. Nature and frequency of integration
7. Program quality	a. High-quality programming
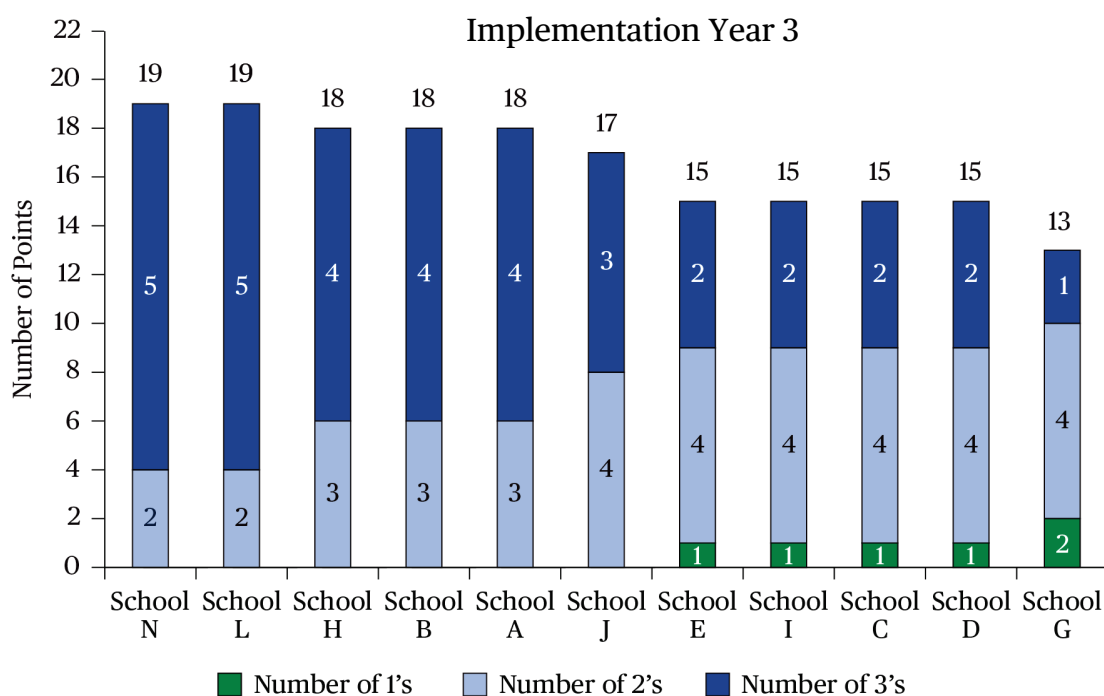
Source: Recreated from Fountain et al. (2016) (pp. 48-49).

Notes: In this example, each school receives a score of 1, 2, or 3, corresponding to ratings of limited, moderate, or full implementation. Thus, the lowest possible score a school could receive in any implementation would be seven and the highest possible score would be 21. Fountain et al. reported the following findings from this figure: "Ratings for the 11 schools in their third year ranged from 13 to 19... Interestingly, most campuses' ratings remained in the 'moderate' category, and instances of 'limited' ratings generally faded as implementation matured—and as the schools that had struggled with various programmatic elements exited the network."

Analytic and reporting approaches for studying implementation context and its variation (research question 2)

The second research question focuses on understanding the implementation context and the degree to which implementation context varied, for example, across study sites or over time. Section III documents the important data needed to understand context. Potential implementers can use these data to better assess whether a specific intervention could be effective in their specific settings. They can do this, in part, by comparing robust information about the context for the study with their own situation, considering how context might have affected study findings and might affect future implementation and outcomes. Reports need to examine and succinctly summarize important contextual information in a visible way. Exhibit 12 shows an example of how Quint et al. (2015) highlighted two important features of context in the report's executive summary of the Success for All (SFA) scale-up demonstration.

Exhibit 12. Example of reporting about context

> ## Context for the evaluation
>
> It is useful to consider the economic and instructional contexts in which the SFA scale-up demonstration has unfolded. These contexts provide a framework through which to view the participating schools' ability to implement the full program model and SFAF's ability to meet its ambitious expansion goals. They also help to define the "counterfactual"–what happens in the absence of the program. Only to the extent that SFA differs from the counterfactual is the program likely to produce impacts. Two trends are worth noting:
>
> - **The effects of the Great Recession and its aftermath.** At the point that SFAF was recruiting schools for the i3 scale-up, many schools and districts were trying to restore positions and services that had been cut as a result of the recession. Furthermore, principals felt that they had less discretion in spending their schools' allocations than had been the case in the past. These circumstances added a new dimension to the challenges already associated with selecting and implementing a new and demanding reading program in high-poverty schools.
>
> - **Heightened focus on reading instruction.** Over the period since SFA was first developed in 1987, reading instruction in the United States has changed markedly. For example, the influence of the National Reading Panel report of 2000, the passage of No Child Left Behind in 2001 (and, as a result, the creation of the Reading First program and the advent of high-stakes testing for grades 3 through 8), the rise of Response to Intervention reading support strategies, and the introduction of the Common Core standards have all contributed to an increased emphasis on phonics and additional interventions for struggling readers. These developments have had the effect of narrowing the differences between schools adopting SFA and schools using other reading programs and have made it harder than it used to be for SFA to "beat the competition."

Source: Recreated from Quint et al. (2015, p. ES-3).

In addition to highlighting key context features in the executive summary, information about context can be presented in other sections of the report including text, tables, figures, or other formats. Tables and figures could relate two or more contextual variables to each other or a context feature to implementation fidelity.

Analytic and reporting approaches for studying the intervention contrast and its variation (research question 3)

Intervention contrast–the focus of research question three–can vary during the course of a study. For example, dimensions of variation can include: (1) multiple sites in a multisite trial; (2) the passage of time in the case of an intervention that spans several months or years; (3) multiple interventions in a multi-arm trial; or (4) some combination of these factors or others. This research question is descriptive in nature and can draw upon both quantitative and qualitative data. Addressing it requires a clear understanding of an intervention's direct and support components as well as the counterfactual condition.

Typically, an analysis of intervention contrast begins with the description of the services that were delivered as part of an intervention and how they compared with services received by students in the counterfactual condition. Exhibit 13 shows how the CUNY ASAP study divided intervention components into four broad categories (for example, requirements, student services) and describes services in each category for ASAP and non-ASAP students.

Exhibit 13. Example of reporting of intervention contrast from the CUNY ASAP study

Key differences between ASAP and usual college services: Three-year impacts report	
ASAP	**Usual college services**
Requirements and messages	
• **Full-time enrollment:** Required • **Taking developmental courses early:** Encouraged consistently and strongly • **Graduating within three years:** Encouraged consistently and strongly	• **Full-time enrollment:** Not required • **Taking developmental courses early:** Encouraged often but not strongly • **Graduating within three years:** Not typically encouraged
Student services	
• **Advising:** Student-to-adviser ratio between 60:1 and 80:1; 95 percent of students met with an adviser during first year and students met with an adviser an average of 38 times in that period • **Career services:** 80 percent of students met with career and employment services staff during first year and students met with such staff an average of 9 times in that period • **Tutoring:** 74 percent of students received tutoring outside of class during first year and students met with a tutor an average of 24 times in that period	• **Advising:** Student-to-adviser ratio between 600:1 and 1,500:1; 80 percent of students met with an adviser during first year and students met with an adviser an average of 6 times in that period • **Career services:** 29 percent of students met with career and employment services staff during first year and students met with such staff an average of 2 times in that period • **Tutoring:** 39 percent of students received tutoring outside of class during first year and students met with a tutor an average of 7 times in that period
Course enrollment	
• **Blocked or linked courses:** Available for first year; few students took complete block of courses, but most students took at least 1 class with a concentration of ASAP students • **ASAP seminar:** Most students took an ASAP seminar for 3 semesters	• **Blocked or linked courses:** Available at 2 colleges during first semester; participation in blocked or linked courses unknown • **Support seminars:** Some students took a freshman seminar or student success course during first year
Financial supports	
• **Tuition waiver:** 3-11 percent of students received waiver, depending on semester • **Free MetroCards:** Most students received free MetroCards for use on public transportation, contingent on participation in the program • **Free use of textbooks:** Most or all students received textbooks	• **Tuition waiver:** Not available • **Free MetroCards:** Not available • **Free use of textbooks:** Not available

Source: MDRC field research data and MDRC student survey, Recreated from Scrivener et al. (2015, p. ES-4).

Study teams can develop useful quantitative measures of intervention contrast. As shown in Exhibit 1 and discussed in Section IV, intervention contrast is the difference in services received by members of the intervention and control groups. Exhibit 14 shows how Millenky et al. (2019) measured these differences in a study of the PACE Center for Girls—a nonresidential year-round program that provides academic and social services to girls ages 11 to 18. This exhibit shows that there was a statistically significant difference between the intervention and control group in students' overall receipt of four out of the five services examined. For example, students in the intervention group were 5 percentage points more likely to enroll in a school or education program in the prior year, and 11 percentage points more likely to receive academic advising services, than their peers in the control group.

Exhibit 14. Example of reporting about contrast

One-year differences in receipt of key services, since random assignment
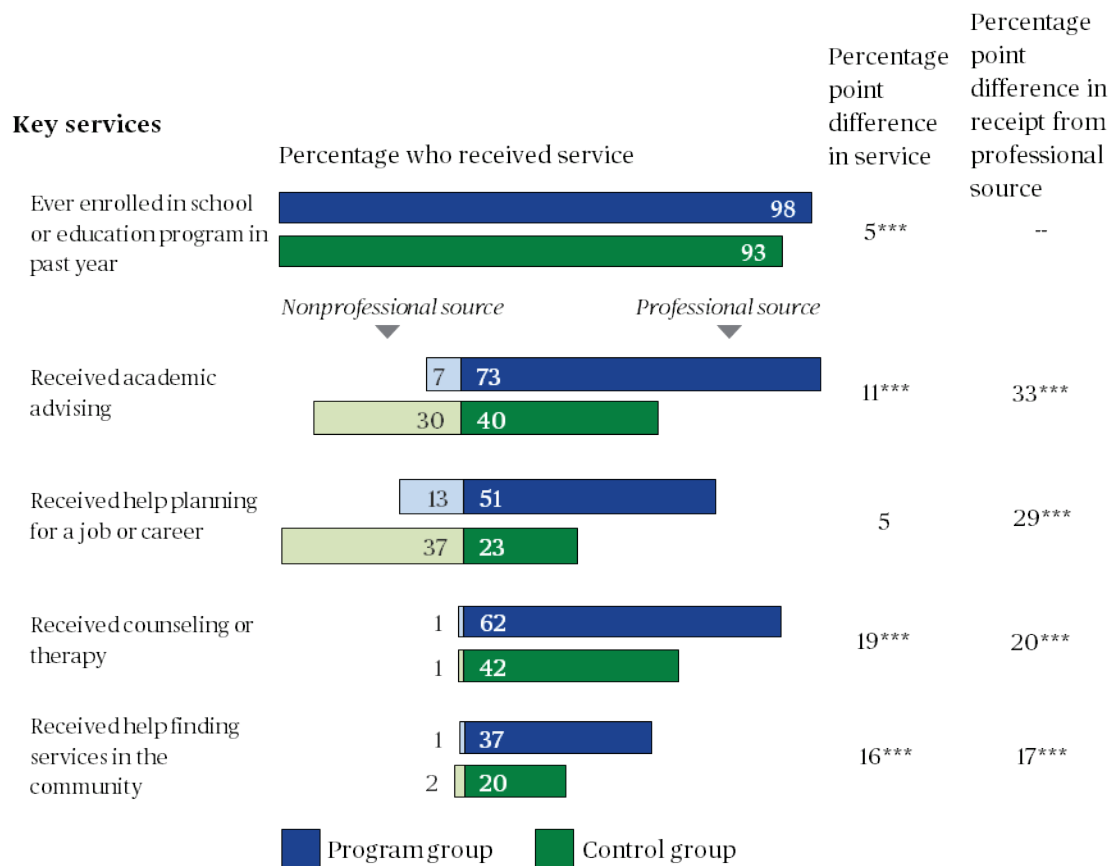
Source: Recreated from Millenky et al. (2019, p. ES-6). MDRC calculations based on response to the PACE evaluation 12-month follow-up survey.

Notes: Results in this figure are regression-adjusted, controlling for pre-random assignment characteristics. Statistical significance levels are indicated as follows: *** = 1 percent; ** = 5 percent; * = 10 percent. Rounding may cause slight discrepancies in sums and differences.

The authors further disaggregate four of the services by provider, distinguishing between whether a service was provided by a non-professional source or a professional source. For example, Exhibit 14 shows that 73 percent of students in the intervention group received academic advising services from a professional source, compared with 40 percent of students in the control group (an intervention contrast of 33 percentage points, which is statistically significant). Whether the contrasts are substantively or practically meaningful is a matter for the entire study team, including the intervention's designers and the sponsoring organization, to consider using reference points such as contrasts measured in similar prior studies or the

control group level on the measure.[23] Whether intervention contrasts are associated with the intervention's impact is the focus of the next section.

Analytic and reporting approaches for studying how intervention components, context, or contrast are associated with impacts (research question 4)

Some studies seek to examine how intervention implementation, context, and contrast are associated with impacts (research question 4). For example, a study team may merge survey, interview, and student outcomes data to explore associations between dosage of services received by participants, intervention contrast, and impacts on student outcomes.

As another example, many education interventions serve diverse students whose backgrounds, social context, and institutional structures can affect their experiences. The study team should consider whether it is feasible to examine differences in the ways that different groups of students experience an intervention as well as how the context and contrast varies across these groups. Whether and how to define the relevant student groups depends on the number and characteristics of students in the study, the subgroups that are the focus of the impact analysis, and the study resources available.

Study teams, intervention developers, and funders often focus on associations between intervention fidelity and impacts. Indeed, linking fidelity to outcomes is the fifth step listed by Nelson et al. (2012) in their five-step model for assessing intervention fidelity. While it may seem intuitive to expect that greater intervention fidelity will be positively associated with improved outcomes, this is not always the case. As shown in Exhibit 1, interventions that are implemented with high fidelity may not be associated with impacts if, for example, there is minimal intervention contrast or the theory embedded in the intervention's logic model is incorrect. For example, the Gray et al. (2022) study of Zoology One compared impacts on literacy outcomes between students in classrooms with high and low intervention fidelity, and found statistically significant differences for some outcomes but not others, concluding "These findings suggest that the intervention is effective on some outcomes even with lower fidelity, while other outcomes require faithful implementation" (p. 17).

Even in impact studies with many sites, determining how intervention implementation, context, and contrast are causally related to impacts can be difficult unless the study is explicitly designed to do so. It might be possible to randomly assign students (or classrooms or schools) to one of several different interventions, each of which involve different components or implementation approaches. But it may not be possible to control variation in

---

[23] Sources that discuss strategies for interpreting substantive significance of effect sizes include Hill et al. (2008), Lipsey et al. (2012), and Kraft (2020). In addition, interpreting substantive significance of contrast estimates can draw on general principles and strategies for writing about data, interpreting substantive significance, and describing quantitative comparisons (for example, Miller 2013, 2015).

context features. More commonly, study teams examine nonexperimental associations between implementation, context, or contrast and impacts. These associations may point to promising implementation approaches or moderating factors that can be explored in further research.

The sources listed in Box 8 provide guidance about methods for exploring nonexperimental associations. Appendix C discusses some standard considerations for focusing analyses on associations that are most relevant and appropriate.

Your study can report findings for research question 4 through tables showing estimated associations.[24] Graphical figures are also useful ways to display the association between impacts and some dimension of implementation. For example, the figure presented in Exhibit 15 shows how Fountain et al.'s (2016) evaluation of Citizen Schools ELT described associations between intervention dosage and impacts on student test scores in English language arts (ELA) and math.

> **Box 8. Resources on nonexperimental analyses of impacts**
>
> These resources from the National Center for Education Evaluation and Regional Assistance provide guidance on and examples of descriptive analyses that can shed light on findings from impact estimates:
>
> *Understanding Variation in Treatment Effects in Education Impact Evaluations: An Overview of Quantitative Methods* (Schochet et al., 2014).
> https://ies.ed.gov/ncee/pubs/20144017/pdf/20144017.pdf.
>
> *Descriptive Analysis in Education: A Guide for Researchers* (Loeb et al., 2017). (Chapter 2 discusses descriptive analyses to support causal understanding.)
> https://ies.ed.gov/ncee/pubs/20174023/pdf/20174023.pdf.

Beyond showing associations between dosage and impacts, Fountain et al. (2016) also examined whether a broad measure of "higher" or "lower" levels of implementation of the intervention was associated with impacts on student test scores. Exhibit 16 shows how these authors presented this analysis in a figure, how they interpreted their findings for readers, and how they interpreted the sensitivity of the findings.

---

[24] For example, see Bloom et al. (2003, p. 564) or Miller et al. (2018, pp. 49-59).

Exhibit 15. Example of reporting associations between intervention quantity (dosage) and impacts

Impact of Citizen Schools ELT on student achievement by dosage

Schools that offer Citizen Schools ELT programming to more than one grade can serve students over multiple years, which may result in cumulative achievement impacts. This possibility is explored in Exhibit 4.9, which indicates that across both outcomes, although the impact estimates were not statistically significant, effect sizes were larger for students who had the potential to have two or more years of Citizen Schools ELT exposure, compared to those with the potential for a single year of dosage.

Exhibit 4.9. Impact of Citizen Schools ELT on ELA and math test scores, by dosage

	One Year	Two or More Years
Number of CS ELT Schools	27	8
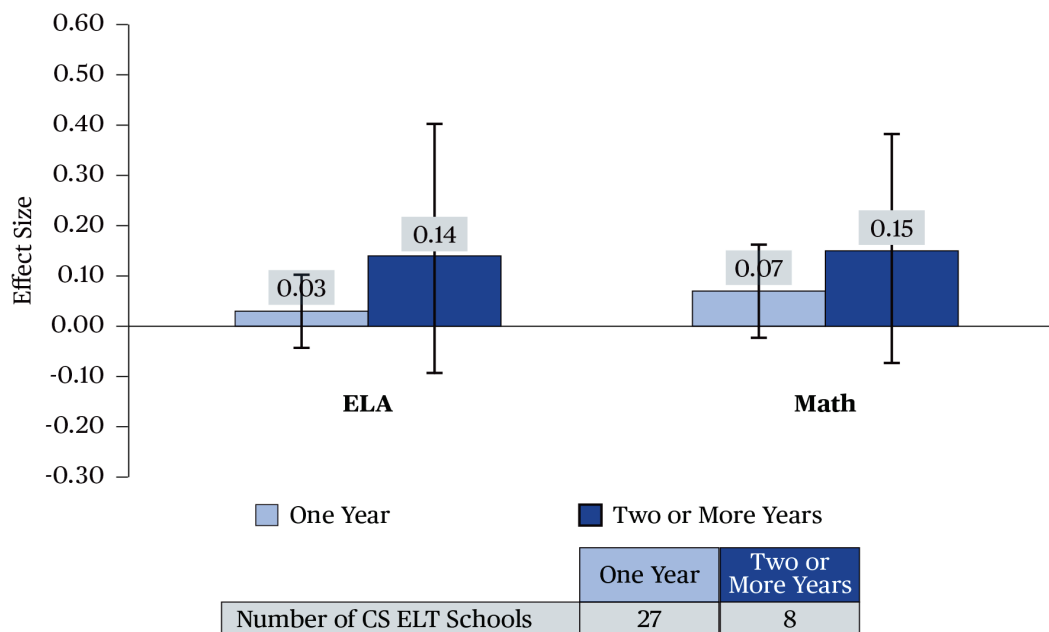
Exhibit reads: After one year of exposure to the Citizen Schools ELT program, students in Citizen Schools ELT schools, on average, scored 0.03 standard deviations higher on their state ELA test than would have been expected in the absence of ELT. The estimated impact was not statistically significant. The upper confidence limit of the impact is 0.11, the lower confidence limit is -0.05.

Source: Publicly available school-level achievement data from states with schools in the Citizen Schools ELT Network.

Notes: + Marginally significant at the $p < 0.10$ level
  * Statistically significant at the $p < 0.05$ level
  ** Statistically significant at the $p < 0.01$ level
  *** Statistically significant at the $p < 0.001$ level

Sample: Standardized ELA test scores from students in 27 ELT schools who received one year of Citizen Schools ELT and their 94 matched comparison schools; standardized ELA test scores from students in 8 ELT schools who received two years or more years of Citizen Schools ELT and their 27 matched comparison schools. Standardized math test scores from students in 27 ELT schools who received one year of Citizen Schools ELT and their 87 matched comparison schools; standardized math test scores from students in 8 ELT schools who received up to two years of Citizen Schools ELT and their 23 matched comparison schools.

Source: Recreated from Fountain et al. (2016, pp. 73-74).

Exhibit 16. Example of reporting associations between intervention implementation and impacts

Impact of Citizen Schools ELT on student achievement by implementation level

As discussed in Chapter 3, schools implement the Citizen School model with varying degrees of fidelity to the intended program. The study also explored whether this variability affects student outcomes by using the implementation index described in the previous chapter to identify "higher" and "lower" implementing Citizen School campuses.[*] Unlike the bar charts presented earlier in the chapter, the bars in Exhibit 4.10 below represent the *difference* in impact between higher and lower implementing schools. Although not statistically significant, students in higher implementing schools score 0.07 and 0.05 standard deviations higher on their ELA and math achievement tests respectively than lower implementing schools.

Exhibit 4.10. Difference in impact of Citizen Schools ELT on ELA and math test scores of higher and lower implementing schools

| Number of CS ELT Higher Implementing Schools | 14 |
| Number of CS ELT Lower Implementing Schools | 12 |
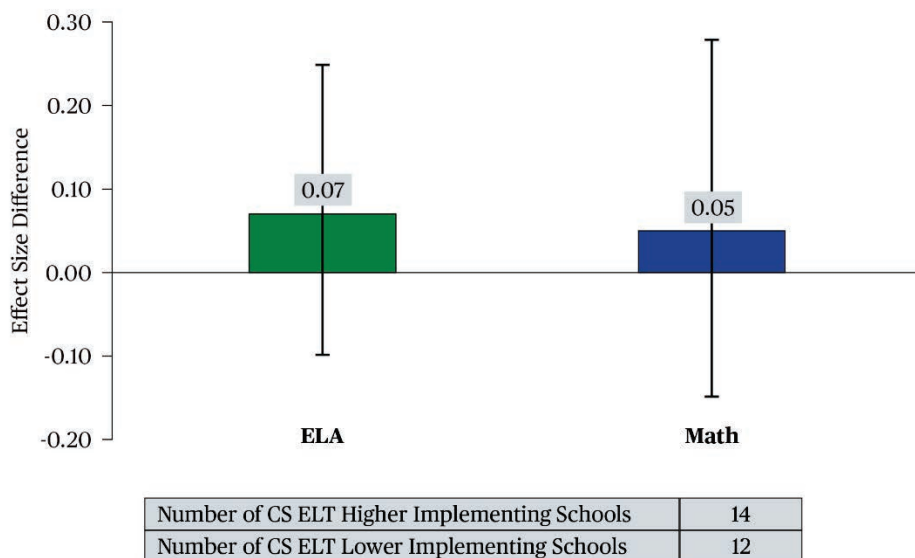
Exhibit reads: The impact on students in higher implementing Citizen Schools ELT schools on their state ELA test is, on average, 0.07 standard deviations higher than the impact on students in lower implementing Citizen Schools ELT schools. The difference between the high and low implementing group was not statistically significant. The upper confidence limit of the difference of the impact is 0.25, the lower confidence limit is -0.10.

[*] For the purposes of these analyses the study team relied solely on interview data, which were collected in schools' first, second, third, and fourth years of implementation (where applicable), whereas survey data were obtained only in schools' second, third, and fourth years. Schools were categorized as high or low implementing schools at the median value of the implementation index total score. Sensitivity analyses assessed whether "high" or "low" designation would change if the index included both interview and survey/quantitative data, and determined that the interview data alone did not yield materially different implementation level designations than the combined index scores.

Source: Recreated from Fountain et al. (2016, p. 73).

Succinct narrative summaries of findings in an executive summary help to elevate findings about associations of intervention implementation, context, and contrast with impacts and to make these findings easily digestible to a broad audience, including practitioners. For example, Exhibit 17 shows how Furgeson et al. (2012) summarized their findings regarding practices of charter school management organizations (CMOs) and the associated impacts.

Exhibit 17. Example of reporting a summary about associations between intervention components and impacts

> **E. Practices associated with positive impacts**
>
> Understanding which CMO practices are associated with the largest impacts can help identify potentially promising educational strategies. To be sure, the associations we observed between impacts and specific CMO practices might not indicate a causal effect of the practices. It is possible that a practice that is positively associated with impacts may in fact be correlated with some other practices we do not observe that are the real driver of student outcomes. But examining associations of practices with impacts is the necessary first step toward identifying promising practices.
>
> **1. Comprehensive behavior policies in schools are associated with larger CMO impacts**
>
> Student impacts in math and reading are larger in CMOs whose schools have comprehensive behavior policies. We found positive associations between student impacts and multiple measures of school behavior policies: Consistent behavior standards and disciplinary policies within a school, zero tolerance policies for potentially dangerous behaviors, behavior codes with student rewards and sanctions, and responsibility agreements signed by students or parents.
>
> **2. CMOs with intensive coaching of teachers tend to have larger positive impacts on student achievement**
>
> Student impacts in math and reading are larger in CMOs with schools that place a greater emphasis on intensive coaching of new teachers. Impacts are associated with a composite measure of teacher coaching that captures the frequency with which teachers are observed and the frequency with which they receive feedback on their performance and their lesson plans. In addition, impacts are larger in those CMOs providing substantial professional development support to their schools.
>
> **3. Several other notable CMO characteristics do not show significant relationships with impacts**
>
> We found no significant relationship between impacts and three other factors that we posited might contribute to student achievement. Specifically, impacts are not correlated with (1) the extent to which CMOs define a consistent educational approach through the selection of curricula and instructional materials, (2) performance-based teacher compensation, or (3) frequent formative student assessments (although impacts are larger when teachers frequently *use* student test results to modify lesson plans). Nor are impacts significantly associated with school or class sizes.
>
> Math impacts are positively correlated with more hours of annual instruction, but this relationship appears to be largely due to the association of instructional time with behavior policies and coaching. We ran multivariate regressions of impacts on key practices that were significantly associated with impacts in bivariate regressions. In the multivariate regressions, the association between impacts and instructional time declined substantially and became statistically insignificant.

Source: Recreated from Furgeson et al. (2012, p. xxxi).

Report findings in ways that support their use and future research

Comprehensive and succinct reporting of details about intervention components, context, and contrast, such as the examples just described, can be helpful to researchers who seek to develop syntheses or meta-analyses. At least one published source from your study should have comprehensive documentation so that other study publications can reference it and other researchers can access it easily. For guidance on the types of information to report, follow reporting guidelines that are required or recommended by standards-setting groups in relevant policy domains or methods groups, by journals or publishers, or by the study's funders. Appendix D includes examples of guidance available from federal agencies and research teams.

As a final note regarding reporting, researchers should always try to convey findings clearly and in ways that make them easier to use. While practitioners and advocates may be most interested in high-level findings and their implications, researchers may be interested in the summary findings as well as various methodological details. One strategy involves developing different products for different audiences, tailoring each product with attention to content, access, and medium (Krause, 2022). Another strategy involves a short summary report with appendices containing other details of interest to specific groups (such as appendices with more details on methods or ones describing the composition of advisory groups and how they participated in the study).

Conclusion

As described in the introduction, this guide focuses on a particular type of implementation research for a particular form of program evaluation: **conducting implementation research as part of high-quality impact studies of education interventions**. It encourages systematic planning for collecting, analyzing, and reporting information about intervention implementation, context, and contrast in these studies, as well as how they are associated with the intervention's impacts. Together, a study's analyses of these topics can help explain why an intervention does or does not produce impacts on specific outcomes.

These explanations and details are essential for closing the gap between education researchers and implementers who want to use research evidence. Whether a specific intervention could be effective in specific settings–that is, understanding the degree of correspondence or fit between the intervention as tested and potential implementers' own situations–is emphasized in frameworks such as Getting to Outcomes® (RAND & the University of South Carolina, 2023), and scholarship on scale-up of evidence-based policies (List et al., 2021) and research use (Supplee et al., 2023). Through the steps described in this guide, researchers can help ensure that information from their studies is accessible and usable to potential implementers of research.

In addition to supporting nuanced understanding of a single study's findings, the suggestions in this guide position studies to contribute to evidence building when combined with other studies in meta-analyses or research syntheses (Cook, 1994; Cooper et al., 2019; The Campbell Collaboration, 2023).

Whether providing potential implementers with a clearer picture of the evidence produced by a single study or contributing to research syntheses, robust implementation research in the context of impact studies is essential for building a strong evidence base about what works, for whom, under what conditions.

References

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*, 4-23. https://doi.org/10.1007/s10488-010-0327-7.

Agodini, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010). *Achievement effects of four early elementary school math curricula: Findings for first and second graders* (NCEE 2011-4001). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/pubs/20094052/pdf/20094052.pdf.

Albrecht, L., Archibald, M., Arseneau, D., & Scott, S. D. (2013). Development of a checklist to assess the quality of reporting of knowledge translation interventions using the Workgroup for Intervention Development and Evaluation Research (WIDER) recommendations. *Implementation Science, 8,* Article 52. https://doi.org/10.1186/1748-5908-8-52.

American Psychological Association. (2021). Inclusive language guidelines. https://www.apa.org/about/apa/equity-diversity-inclusion/language-guidelines.

Bauer, M. S., & Kirchner, J. (2020). Implementation science: What is it and why should I care? *Psychiatry Research*, *283*. https://doi.org/10.1016/j.psychres.2019.04.025.

Bloom, H. S., Hill, C. J., & Riccio, J. A. (2003). Linking program implementation and effectiveness: Lessons from a pooled sample of welfare-to-work experiments. *Journal of Policy Analysis and Management*, *22*(4), 551-575. https://doi.org/10.1002/pam.10154.

Boulay, B., Goodson, B., Olsen, R., McCormick, R., Darrow, C., Frye, M., Gan, K., Harvill, H., & Sarna, M. (2018). *The Investing in Innovation Fund: Summary of 67 evaluations: Final report* (NCEE 2018-4013). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/pubs/20184013/pdf/20184013.pdf.

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics*, *130*(5), e1136-e1145. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3483890.

Brown, C. H., Curran, G., Palinkas, L. A., Aarons, G. A., Wells, K. B., Jones, L., Collins, L. M., Duan, N., Mittman, B. S., Wallace, A., Tabak, R. G., Ducharme, L., Chambers, D. A., Neta, G., Wiley, T., Landsverk, J., Cheung, K., & Cruden, G. (2017). An overview of research and evaluation designs for dissemination and implementation. *Annual Review of Public Health*, *38,* 1-22. https://doi.org/10.1146/annurev-publhealth-031816-044215.

Cerna, O., & Condliffe, B. (2021). *Guiding questions for supporting culturally responsive evaluation practices and an equity-based perspective.* MDRC. https://www.mdrc.org/sites/default/files/Equity-Guiding_Questions.pdf.

Clements, D. H., Agodini, R., & Harris, B. (2013). *Instructional practices and student math achievement: Correlations from a study of math curricula. NCEE Evaluation Brief.* (NCEE 2013-4020). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/pubs/20134020/.

Cohen, D., & Crabtree, B. (2006). *Qualitative research guidelines project.* The Robert Wood Johnson Foundation. http://www.qualres.org/index.html.

Cook, C. R., Lyon, A. R., Locke, J., Waltz, T., & Powell, B. J. (2019). Adapting a compilation of implementation strategies to advance school-based implementation research and practice. *Prevention Science, 20*(6), 914-935. https://doi.org/10.1007/s11121-019-01017-1.

Cook, T. D. (Ed.) (1994). *Meta-analysis for explanation: A case book.* Russell Sage Foundation.

Cooper, H., Hedges, L. V., & Valentine J. C. (Eds.) (2019). *The handbook of research synthesis and meta-analysis* (3rd ed.). Russell Sage Foundation.

Cordray, D. S., & Pion, G. M. (2006). Treatment strength and integrity: Models and methods. In R. R. Bootzin & P. E. McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation* (pp. 103-124). American Psychological Association.

Corrin, W., Sepanik, S., Gray, A., Fernandez, F., Briggs, A., & Wang, K. (2014). *Laying tracks to graduation: The first year of implementing Diplomas Now.* MDRC. https://www.mdrc.org/sites/default/files/Diplomas_Now_First_Year_FR_0.pdf.

Creswell, J. W. (2021). *A concise introduction to mixed methods research* (2nd ed.). SAGE Publications.

Creswell, J. W. & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Curran, G. M. (2020). Implementation science made too simple: a teaching tool. *Implementation Science Communications, 1*(1), 27. https://doi.org/10.1186/s43058-020-00001-z.

Curran, G. M., Bauer, M., Mittman, B., Pyne, J. M., & Stetler, C. (2012). Effectiveness-implementation hybrid designs. *Medical Care, 50*(3), 217-226. https://doi.org/10.1097/MLR.0b013e3182408812.

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science, 4*(1), 1-15. https://doi.org/10.1186/1748-5908-4-50.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*(1), 23-45. https://doi.org/10.1016/S0272-7358(97)00043-3.

Dhillon, S., Darrow, C., & Meyers, C. V. (2015). Introduction to implementation fidelity. In C. V. Meyers & W. C. Brandt, W. C. (Eds.), *Implementation fidelity in education research: Designer and evaluator considerations* (pp. 8-22). Routledge.

Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K., Buckley, J. A., Olin, S., Romanelli, L. H., Leaf, P. J., Greenberg, M. T., & Ialongo, N. S. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion*, *1*(3), 6-28. https://www.tandfonline.com/doi/abs/10.1080/1754730X.2008.9715730.

Dong, N., Kelcey, B., & Spybrook, J. (2020). Design considerations in multisite randomized trials probing moderated treatment effects. *Journal of Educational and Behavioral Statistics*, *46*(5), 527-559. https://doi.org/10.3102/1076998620961492.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3-4), 327-350. https://doi.org/10.1007/s10464-008-9165-0.

Dymnicki, A., Trivits, L., Hoffman, C. & Osher, D. (2020). *Advancing the use of core components of effective programs: Suggestions for researchers publishing evaluation results*. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. https://aspe.hhs.gov/sites/default/files/private/pdf/263931/ASPE-Brief-Core-Components.pdf.

EdInstruments. (2023). *Instruments: School climate*. March 30. https://edinstruments.com/instruments/domains/5/subdomains/34.

Education Commission of the States. (2023). https://www.ecs.org/.

Federal Reserve Economic Data. (2023). *Federal Reserve Economic Data (FRED)*. https://fred.stlouisfed.org/.

Finch, W. H. (2020). *Exploratory factor analysis*. SAGE Publications

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. National Implementation Research Network, Louis de la Parte Florida Mental Health Institute, University of South Florida.

Fountain, A. R, Gamse, B., Velez, M., Hillard, M., & Cropper, P. (2016). *Evaluation of Citizen Schools' expanded learning time model: Final report*. Abt Associates. https://www.abtassociates.com/sites/default/files/2019-12/Citizen%20Schools%20ELT%20Evaluation%20Final%20Report.pdf.

Frank Porter Graham Child Development Institute. (n.d.). *Early Childhood Environment Rating Scale, Third Edition*. University of North Carolina at Chapel Hill. https://ers.fpg.unc.edu/scales-early-childhood-environment-rating-scale-third-edition.

Furgeson, J., Gill, B., Haimson, J., Killewald, A., McCullough, M., Nichols-Barrer, I., Teh, B., & Verbitsky-Savitz, N. (2012). *Charter-school management organizations: Diverse strategies and diverse student impacts*. Mathematica Policy Research. https://mathematica.org/publications/charterschool-management-organizations-diverse-strategies-and-diverse-student-impacts.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Goodson, B., Darrow, C., Wolf, A., Price, C., & Boulay, B. (2019). *Guidance for planning and reporting your Education Innovation and Research (EIR) fidelity of implementation study* (Prepared for the Institute of Education Sciences, U.S. Department of Education). Abt Associates, Inc.

Gray, A. M., Sirinides, P. M., Fink, R. E., & Bowden, A. B. (2022). Integrating literacy and science instruction in kindergarten: Results from the efficacy study of *Zoology One*. *Journal of Research on Educational Effectiveness*, *15*(1), 1-27. https://doi.org/10.1080/19345747.2021.1938313.

Hamilton, G., & Scrivener, S. (2018). Measuring treatment contrast in randomized controlled trials (MDRC Working Paper.) https://www.mdrc.org/sites/default/files/MTC_Paper_MDRC_WEBSITE_VERSION.pdf.

Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools? *Exceptional Children*, *79*(2), 181-193. https://doi.org/10.1177/001440291307900205.

Herman, K. C., Dong, N., Reinke, W. M., & Bradshaw, C. P. (2022). Accounting for traumatic historical events in educational randomized controlled trials. *School Psychology Review*, 1-17. https://doi.org/10.1080/2372966X.2021.2024768.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172-177. https://doi.org/10.1111/j.1750-8606.2008.00061.x.

Hill, H. C., & Erickson, A. (2019). Using implementation fidelity to aid in interpreting program impacts: A brief review. *Educational Researcher*, *48*(9), 590-598. https://doi.org/10.3102/0013189X19891436.

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Woods, M., McCulloch, P., Wyatt, J. C., Chan, A. W., & Michie, S. (2014). Better reporting of interventions: Template for Intervention Description and Replication (TIDieR) checklist and guide. *BMJ*, *348*. https://doi.org/10.1136/bmj.g1687.

Horner, R. H., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. B. (2004). The School-wide Evaluation Tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Interventions*, *6*(1), 3-12. https://journals.sagepub.com/doi/10.1177/10983007040060010201.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*(1), 88-110. https://doi.org/10.1080/19345740802539325.

Hulleman, C. S., Rimm-Kaufman, S. E., & Abry, T. (2013). Innovative methodologies to explore implementation: Whole-part-whole–Construct validity, measurement, and analytical issues for intervention fidelity assessment in education research. In T. Halle, A. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 65-93). Paul H Brookes Publishing Co.

Institute of Behavioral Research. (2009). *TCU Organizational Readiness for Change (ORC-D4)*. Texas Christian University, Institute of Behavioral Research. https://ibr.tcu.edu/forms/organizational-staff-assessments/.

Institute of Education Sciences. (2021a). *Document treatment implementation and contrast*. https://ies.ed.gov/seer/implementation.asp.

Institute of Education Sciences. (2021b). *What Works Clearinghouse (WWC) reporting guide for study authors*. U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/WWC_Author_Guide_Jul2021.pdf.

Institute of Education Sciences. (2022a, June 22). *Standards for excellence in education research*. U.S. Department of Education. https://ies.ed.gov/seer/index.asp.

Institute of Education Sciences. (2022b). *Equity*. https://ies.ed.gov/seer/equity.asp.

Institute of Education Sciences. (2023a). *Common core of data*. https://nces.ed.gov/ccd/files.asp.

Institute of Education Sciences. (2023b). *Education demographic and geographic estimates*. https://nces.ed.gov/Programs/EDGE.

Jacob, R., Doolittle, F., Kemple, J., & Somers, M. (2019). A framework for learning from null results. *Educational Researcher*, *48*(9), 580-589. https://doi.org/10.3102/0013189X19891955.

Kelcey, B., Dong, N., Spybrook, J., & Shen, Z. (2017). Experimental power for indirect effects in group-randomized studies with group-level mediators. *Multivariate Behavioral Research*, *52*(6), 699-719. https://doi.org/10.1080/00273171.2017.1356212.

Kelcey, B., Spybrook, J., Dong, N., & Bai, F. (2020). Cross-level mediation in school-randomized studies of teacher development: Experimental design and power. *Journal of Research on Educational Effectiveness*, *13*(3), 459-487. https://doi.org/10.1080/19345747.2020.1726540.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241-253. https://doi.org/10.3102/0013189X20912798.

Krause, H. (2022, August 12). *The two report strategy. We All Count*. https://weallcount.com/2022/08/12/the-two-report-strategy/.

Lemons, C. J., Fuchs, D., Gilbert, J. K., & Fuchs, L. S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher*, *43*(5), 242-252. https://doi.org/10.3102/0013189X14539189.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncser/pubs/20133000/pdf/20133000.pdf.

List, J. A., Suskind, D., & Suplee, L. H. (eds.) (2021). *The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it.* New York: Routledge.

Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). *Descriptive analysis in education: A guide for researchers* (NCEE 2017-4023). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/pubs/20174023/pdf/20174023.pdf.

Lyon, A. R., & Bruns, E. J. (2019). User-centered redesign of evidence-based psychosocial interventions to enhance implementation–hospitable soil or better seeds? *JAMA Psychiatry, 761*(1), 3-4. https://doi.org/10.1001/jamapsychiatry.2018.3060.

Maier, M. F., Hsueh, J., & McCormick, M. (2020). *Rethinking classroom quality: What we know and what we are learning.* MDRC. https://www.mdrc.org/sites/default/files/Conceptualization_Classroom_Quality_0.pdf.

Maier, M., & Mattera, S. (2015). *Making Pre-K count: Implementation measurement and analysis plan.* [Unpublished manuscript]. New York: MDRC.

Manno, M. (2012). *Implementation research analysis plan for the Youth Villages Transitional Living Study.* [Unpublished manuscript]. New York: MDRC.

Manno, M., & Treskon, L. (in press). Applying multiple methods to implementation research: Applications from youth education and employment programs. In A. Rangarajan & D. Paulsell (Eds.), *The Oxford handbook of social program design and implementation evaluation.* Oxford University Press.

Marti, C. N. (2008). Dimensions of student engagement in American community colleges: Using the Community College Student Report in research and practice. *Community College Journal of Research and Practice, 33*(1), 1-24. https://www.tandfonline.com/doi/full/10.1080/10668920701366867.

Mattera, S., Jacob, R., & Morris, P. (2018). *Strengthening children's math skills with enhanced instruction: The impacts of Making Pre-K Count and High 5s on kindergarten outcomes.* MDRC. https://www.mdrc.org/sites/default/files/MPC-High_5s_Impact_FR_0.pdf.

May, H., Sirinides, P. M., Gray, A., & Goldsworthy, H. (2016). *Reading recovery: An evaluation of the four-year i3 scale-up.* CPRE Research Reports. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1089&context=cpre_researchreports.

McClenney, K., Marti, C. N., & Adkins, C. (n.d.). *Student engagement and student outcomes: Key findings from CCSSE validation research*. Community College Survey of Student Engagement. https://www.ccsse.org/aboutsurvey/docs/CCSSE%20Validation%20Summary.pdf.

McCormick, M., & Maier, M. (2018). *Working with practitioners to develop measures of implementation fidelity*. MDRC. https://www.mdrc.org/publication/working-practitioners-develop-measures-implementation-fidelity.

McCoy, D. C., Morris, P. A., Connors, M. C., Gomez, C. J., & Yoshikawa, H. (2016). Differential effectiveness of Head Start in urban and rural communities. *Journal of Applied Developmental Psychology, 43*, 29-42. https://doi.org/10.1016/j.appdev.2015.12.007.

MDRC. (2015). *Accelerated Study in Associate Programs (ASAP): A program that almost doubles three-year graduation rates*. https://www.mdrc.org/publication/accelerated-study-associate-programs-asap.

MDRC (n.d.). *Making Pre-K Count: Project overview*. https://www.mdrc.org/project/making-pre-k-count#overview.

Meyers, C., & Brandt, W. C. (Eds.). (2014). *Implementation fidelity in education research: Designer and evaluator considerations* (1st ed.). Routledge. https://doi.org/10.4324/9781315795089.

Meyers, C., Molefe, A., Dhillon, S., & Zhu, B. (2015). *The impact of eMINTS professional development on teacher instruction and student achievement: Year 3 report.* American Institutes for Research. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.694.9036&rep=rep1&type=pdf.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2019). *Qualitative data analysis: A methods sourcebook* (4th ed.). Sage Publications.

Millenky, M., Treskon, L., Freedman, L., & Mage, C. (2019). *Focusing on girls' futures: Results from the evaluation of PACE Center for Girls*. MDRC. https://www.mdrc.org/publication/focusing-girls-futures.

Miller, C., Cummings, D., Millenky, M., & Wiegand, A. (2018). *Laying a foundation: Four-year results from the national YouthBuild evaluation*. MDRC. https://www.mdrc.org/sites/default/files/YouthBuild_Final_508%20compliant.pdf.

Miller, C., Headlam, C., Manno, M., & Cullinan, D. (2020). *Increasing community college graduation rates with a proven model: Three-year results from the Accelerated Study in Associate Programs (ASAP) Ohio demonstration*. MDRC. https://www.mdrc.org/sites/default/files/ASAP_OH_3yr_Impact_Report_1.pdf.

Miller, J. E. (2013). *The Chicago guide to writing about multivariate analysis* (2nd ed.). The University of Chicago Press.

Miller, J. E. (2015). *The Chicago guide to writing about numbers* (2nd ed.). The University of Chicago Press.

Moullin, J. C., Dickson, K. S., Stadnick, N. A., Rabin, B., & Aarons, G. A. (2019). Systematic review of the Exploration, Preparation, Implementation, Sustainment (EPIS) framework. *Implementation Science, 14,* 1. https://doi.org/10.1186/s13012-018-0842-6.

National Center on Safe Supportive Learning Environments. (2022). *School climate measurement.* American Institutes for Research. https://safesupportivelearning.ed.gov/topic-research/school-climate-measurementl.

Neild, R. C., Robinson, D., & Agufa, J. (2022). *Sharing study data: A guide for education researchers* (NCEE 2022-004). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/pubs/2022004/.

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research*, *39*(4), 374–396. https://doi.org/10.1007/s11414-012-9295-x.

Nilsen, P., & Bernhardsson, S. (2019). Context matters in implementation science: A scoping review of determinant frameworks that describe contextual determinants for implementation outcomes. *BMC Health Services Research*, *19*(1), 1–21. https://doi.org/10.1186/s12913-019-4015-3.

Office of the State Superintendent of Education for the District of Columbia. (2023). *District of Columbia early learning standards.* https://osse.dc.gov/publication/district-columbia-early-learning-standards-dc-els.

Palinkas, L. A., Aarons, G. A., Horwitz, S., Chamberlain, P., Hurlburt, M., & Landsverk, J. (2011). Mixed method designs in implementation research. *Administration and Policy in Mental Health*, *38*(1), 44–53. https://doi.org/10.1007/s10488-010-0314-z.

Patton, M. Q. (2014). *Qualitative research and evaluation methods: Integrating theory and practice* (4th ed.). SAGE Publications.

Pinnock, H., Epiphaniou, E., Sheikh, A., Griffiths, C., Eldridge, S., Craig, P., & Taylor, S. J. C. (2015). Developing standards for reporting implementation studies of complex interventions (StaRI): A systematic review and e-Delphi. *Implementation Science*, *10*(1), 1–10. https://doi.org/10.1186/s13012-015-0235-z.

Powell, B. J., Mettert, K. D., Dorsey, C. N., Weiner, B. J., Stanick, C. F., Lengnick-Hall, R., Ehrhart, M. G., Aarons, G. A., Barwick, M. A., Damschroder, L. J., & Lewis, C. C. (2021). Measures of organizational culture, organizational climate, and implementation climate in behavioral health: A systematic review. *Implementation Research and Practice*, *2*, 1–29. https://doi.org/10.1177/26334895211018862.

Powell, B. J., Waltz, T. J., Chinman, M. J., Damschroder, L. J., Smith, J. L., Matthieu, M. M., Proctor, E. K., & Kirchner, J. E. (2015). A refined compilation of implementation strategies: Results from the Expert Recommendations for Implementing Change (ERIC) project. *Implementation Science*, *10*(21), 1–14. https://doi.org/10.1186/s13012-015-0209-1.

Price, C., Goodson, B., Wolf, A., & Boulay, B. (2016). *Evaluation plan template*. Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/projects/pdf/EvaluationPlanTemplate.pdf.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health*, *38*(2), 65-76. https://doi.org/10.1007/s10488-010-0319-7.

Quint, J., Zhu, P., Balu, R., Rappaport, S., & DeLaurentis, M. (2015). *Scaling up the Success for All model of school reform: Final report from the Investing in Innovation (i3) evaluation.* MDRC. https://www.mdrc.org/sites/default/files/SFA_2015_FR.pdf.

RAND & University of South Carolina. (2023). *Getting to Outcomes®: A proven method for improving behavioral health programs.* https://www.rand.org/health-care/projects/getting-to-outcomes.html.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* SAGE Publications.

Reading Recovery Council of North America. (2012). *Standards and Guidelines of Reading Recovery in the United States,* 6th ed. Reading Recovery Council of North America.

Remler, D. K., & Van Ryzin, G. G. (2021). *Research methods in practice: Strategies for description and causation* (3rd ed.). SAGE Publications.

Roos, J. M., & Bauldry, S. (2021). *Confirmatory factor analysis.* SAGE Publications.

Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2018). *Evaluation: A systematic approach*. Sage Publications.

Sanetti, L. M. H., & Collier-Meek, M. A. (2019). *Supporting successful interventions in schools: Tools to plan, evaluate, and sustain effective implementation.* The Guilford Press.

Schochet, P. Z., Puma, M., & Deke, J. (2014). *Understanding variation in treatment effects in education impact evaluations: An overview of quantitative methods* (NCEE 2014-4017). National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/pubs/20144017/pdf/20144017.pdf.

Scott, V. C., Jillani, Z., Malpert, A., Kolodny-Goetz, J., & Wandersman, A. (2022). A scoping review of the evaluation and effectiveness of technical assistance. *Implementation Science Communications*, *3*(70), 1-16. https://doi.org/10.1186/s43058-022-00314-1.

Scrivener, S., Weiss, M. J., Ratledge, A., Rudd, T., Sommo, C., & Fresques, H. (2015). *Doubling graduation rates: Three-year effects of CUNY's Accelerated Study in Associate Programs (ASAP) for developmental education students.* MDRC. https://www.mdrc.org/sites/default/files/doubling_graduation_rates_fr.pdf.

Shakman, K., & Rodriguez, S. M. (2015). *Logic models for program design, implementation, and evaluation: Workshop toolkit* (REL 2015-057). Regional Educational Laboratory Northeast & Islands, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/rel/regions/northeast/pdf/rel_2015057.pdf.

Society for Implementation Research Collaboration. (2021, January 13). *The SIRC instrument review project (IRP): A systematic review and synthesis of implementation science instruments.* https://societyforimplementationresearchcollaboration.org/sirc-instrument-project/.

Spybrook, J., Zhang, Q., Kelcey, B., & Dong, N. (2020). Learning from cluster randomized trials in education: An assessment of the capacity of studies to determine what works, for whom, and under what conditions. *Educational Evaluation and Policy Analysis*, *42*(3), 354-374. https://doi.org/10.3102/0162373720929018.

Stange, K. C., & Glasgow, R. E. (2013). *Contextual Factors: The importance of considering and reporting context in research on the patient-centered medical home*. Agency for Healthcare Research and Quality. AHRQ Publication No. 13-0045-EF. https://www.ahrq.gov/sites/default/files/wysiwyg/ncepcr/tools/PCMH/contextual-factors.pdf.

Stuart, E. A. (2012). Recent advances in missing data methods: Imputation and weighting. Prepared for the Maternal Child Health Bureau, Health Resources and Services Administration. https://www.mchnavigator.org/documents/Stuart-MissingData.pdf.

Supplee, L., Boaz, A., & Metz, A. (2023). *Learning across contexts: Bringing together research on research use and implementation science*. William T. Grant Foundation. https://wtgrantfoundation.org/digest/learning-across-contexts-bringing-together-research-on-research-use-and-implementation-science.

The Campbell Collaboration. (2023). https://www.campbellcollaboration.org/.

Tipton, E., & Olsen, R. B. (2022). *Enhancing the generalizability of impact studies in education* (NCEE 2022-003). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/pubs/2022003/.

Trainor, A., & Bal, A. (2014). Development and preliminary analysis of a rubric for culturally responsive research. *The Journal of Special Education*, *47*(4), 203-216. https://doi.org/10.1177/0022466912436397.

University of Virginia, School of Education and Human Development. (n.d.). *Classroom Assessment Scoring System (CLASS)™*. https://education.virginia.edu/classroom-assessment-scoring-system.

University of Washington. (2023a). *The UW Implementation Science Resource Hub*. https://impsciuw.org/.

University of Washington. (2023b). *The Research Institute for Implementation Science in Education*. https://smartcenter.uw.edu/programs-services/riise/.

Wagner, H., & Wright, D. (2022). *Understanding qualitative and mixed method approaches: Resources for further reading* (OPRE Report 2023-006). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. https://www.acf.hhs.gov/opre/report/understanding-qualitative-mixed-methods-approaches-resources-further-reading.

Watkins, M. W. (2018). Exploratory factor analysis: a guide to best practice. *Journal of Black Psychology*, *44*(3), 219-246. https://doi.org/10.1177/0095798418771807.

Weiner, B. J. (2009). A theory of organizational readiness for change. *Implementation Science*, *4*(1), 1-9. https://doi.org/10.1186/1748-5908-4-67.

Weiner, B. J., Amick, H., & Lee, S. Y. (2008). Conceptualization and measurement of organizational readiness for change: A review of the literature in health services research and other fields. *Medical Care Research and Review*, *65*(4), 379-436. https://doi.org/10.1177/1077558708317802.

Weiss, M. J., Bloom, H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management*, *33*(3), 778-808. http://www.jstor.org/stable/24033389.

Weiss, M. J., Bloom, H. S., & Singh, K. (2022). What 20 years of MDRC RCTs suggest about predictive relationships between intervention features and intervention impacts for community college students. *Educational Evaluation and Policy Analysis*. https://doi.org/10.3102/01623737221139493.

Wiegand, A., Manno, M., Leshnick, S., Treskon, L., Geckeler, C., Lewis-Charp, H., Sinicrope, C., Clark, M., & Nicholson, B. (2015). Adapting to local context: Findings from the YouthBuild evaluation implementation study. MDRC. https://www.mdrc.org/sites/default/files/Adapting_to_Local_Context.pdf.

Wilson, S. J., Lipsey, M. W., Aloe, A., & Sahni, S. (2020). *Developing evidence-based practice guidelines for youth programs: Technical report on the core components of interventions that address externalizing behavior problems.* Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. https://aspe.hhs.gov/sites/default/files/documents/cf87486557ccdb62fe765f4378568357/technical-report-externalizing.pdf.

Wiltsey Stirman, S., Baumann, A. A., & Miller, C. J. (2019). The FRAME: An expanded framework for reporting adaptations and modifications to evidence-based interventions. *Implementation Science*, *14*, 1-10. https://doi.org/10.1186/s13012-019-0898-y.

Woodson, T. T. (2021). Using a culturally responsive and equitable evaluation approach to guide research and evaluation. Mathematica. https://www.mathematica.org/publications/using-a-culturally-responsive-and-equitable-evaluation-approach-to-guide-research-and-evaluation.

Appendix A. Template for implementation research analysis plan

This appendix provides an analysis plan for the implementation research conducted as part of an impact study, as described in this guide. It contains the following sections:

- Context
- Introduction to the intervention model
- Logic model
- Analysis plan
    - Key research questions
    - Data collection strategy
    - Analytic approach
- Reporting the findings

Context

This section should provide discussion of the broader context and history of the policy issue that the intervention aims to affect (for example, who are the stakeholders and what challenges have they faced that the intervention aims to address? What is the larger social or policy context of the study sites? What other services have study participants historically been able to access?). It can also include information about the context in which the study will unfold, as well as strategies the study team will take to learn about context as part of the implementation research process.

Introduction to the intervention model

This section should provide information about the intervention's direct components and support components. Direct components can be characterized along four dimensions: content (the basic ingredients of the intervention), quantity (how frequently and for what duration each component will occur), mode (how, by whom, and when the intervention will be delivered), and quality (tools to measure quality may depend on the nature of the intervention). See Section II of this guide.

Logic model

Given the complexities of program implementation and the factors that can influence it, state clearly the aspects of the logic model and implementation processes that your study will focus most sharply on and why. It may be helpful to also articulate the elements that you do not plan to devote many resources to, to ensure that all stakeholders have thought carefully about and agree to boundaries of the study and high priority questions (see Section II of guide). See the following example logic model:

Exhibit A.1. Example logic model based on Manno (2012)

Youth supports	TI counselors Family Friends Foster family		Key short-term outcomes (*administrative records and/or 12-month survey*)	Key long-term outcomes (*administrative records and/or 24-month survey*)
Youth who (a) spent at least 365 days in state custody after age 14 (not necessarily continuously) or (b) spent at least one day in custody after age 17.

Excludes youth with history of violent crime, with continued and unremitting involvement in gang activity or serious substance abuse, or those who are not considered able to live independently due to serious mental illness or developmental disability. | **Barriers to successful transition to adulthood**

• Lack of permanent or strong connection to responsible adult(s)
• Housing instability including while in state custody
• Truancy
• Poor educational performance
• Lack of educational continuity
• Limited opportunity to learn about working world
• Domestic violence
• Mental health concerns
• Substance abuse problems
• Poor money management skills
• Risky sexual behavior or unplanned parenthood
• Lack of health insurance
• Involvement with juvenile justice system
• Disconnected from community supports | **Transitional living program components**

• Connection with caring adult
• Implementation of housing plan
• Education attainment/ maintenance
• Employment and job seeking skills
• Management of safe relationships
• Alleviate symptoms of mental health
• Management of substance abuse
• Personal health care and well-being
• Community safety
• Money management
• Connections to community supports
• Financial supports
• Referrals for other services | • Increased positive social support, including relationships with caring adult
• Increased level of stable and/or suitable housing
• Increased high school diploma or GED attainment
• Increased post-secondary education enrollment
• Increased post-secondary education persistence
• *Increased employment*
• Reduced pregnancy
• Increased safety of children (children of participants)
• Reduced risk behaviors (substance abuse, sexual activity, delinquency)
• *Reduced criminal behavior and justice system involvement*
• Decreased victimization (spousal/partner or otherwise)
• *Increased receipt of government supports (TANF/Food Stamps/ Medicaid)* | • Increased positive social support, including relationships with caring adult
• Increased level of stable and/or suitable housing
• *Reduced arrest/ conviction rates*
• Increased high school diploma or GED attainment
• *Increased post-secondary education attainment*
• *Increased employment and earnings*
• Reduced pregnancy
• Increased safety of children (children of participants)
• Decreased victimization
• *Decreased receipt of government supports (TANF/Food Stamps/ Medicaid)* |

National context

- H.R. 6893: Fostering Connections to Success and Increasing Adoptions Act of 2008
- Foster Care Independence Act of 1999, including the John H. Chafee Foster Care Independence Program

State and local context

- Changing state policies for foster care and former foster care youth
- Tennessee Department of Children Services - Independent living services for foster youth
- Private sector entities providing post-custody services
- Court system oversight of youth in state custody
- Service Environments of Tennessee Regions (west, middle, east)
- Service Environments of surrounding states (Arkansas. Mississippi)

Analysis plan

As described in this guide, implementation research can be expansive and cover numerous topics and constructs. This section should articulate what the study will and will not measure and should outline the different ways the study will measure a construct (both qualitatively and quantitatively).

Key research questions

Discuss research questions that will guide implementation research activities. Research questions often address the components of the intervention and how they are implemented, the context in which the intervention is being implemented, the role of contrast, and associations between implementation, context, contrast, and impact (see Section I of this guide).

Data collection strategy

This section should specify details for defining and measuring direct and support components of implementation (see Section II of this guide). It should include:

- Description of measures and data sources
  - Some projects find it useful to present data sources in a matrix alongside the research topics or questions each data source will help answer (see Exhibit A.1).
- Description of data collection methods
  - If teams create a matrix for questions and data sources, a separate column could be added to describe data collection plans for each source.
  - Implementation researchers should keep in mind that all data–both quantitative and qualitative–that pertain to measuring or describing implementation fall under their purview. Data collection plans should clearly articulate how the implementation research team will coordinate with the impact team to access participation data, management information systems (MIS) data, and survey data, for example. If the team plans to conduct observations that require quantitative analysis, plans should clarify who is responsible for this task.
  - Plans should also define what types of data can be collected and how this will be accomplished with consistency across sites.

Exhibit A.2. Example of data sources from evaluation of Youth Villages

This section of the analysis plan may present the information above in a table similar to the example shown here, which includes research areas of interest (you may also wish to include specific research questions) and data sources the study will rely on to address them. Although this example does not include intervention contrast as a research area, you should plan to include it in your description of data collection plans and any relevant tables if your study will address questions related to intervention contrast (See Section IV of this guide).

Research topic	Interviews with staff, manage-ment	Interviews with other stake-holders	Document review	MIS	Baseline or follow-up survey	Adminis-trative records	Interviews or focus groups with youth
Local context and service environment	X	X			X	X	X
Demographic, economic, social forces	X					X	
Other organizations serving youth	X	X			X		X
Organizational structure	X		X				
Youth Villages history	X		X				
Transitional Living structure	X						
Intended intervention	X		X				
Study recruitment	X	X	X				
Intervention implementation	X	X	X	X	X		X
Fidelity to model	X		X				
Service delivery	X			X	X		X
Influence of evaluation	X	X					
Service utilization	X			X	X		X
Participation	X			X	X		X
Satisfaction with services					X		X
Implementation lessons	X	X	X	X	X		X

Source: Manno (2012).

Analytic approach

This section should describe your analytic strategy for quantitative and qualitative implementation data, as well as any approaches you will take to having implementation data speak to impact data (see Section V of this guide). Some things to highlight include the following:

- To what extent will your quantitative and qualitative implementation data complement each other? For instance, a team may primarily use qualitative data to develop themes, then identify areas of overlap with quantitative data and do selective analysis to help fill out their story. Conversely, the team could begin by analyzing quantitative data and let those findings drive the qualitative analysis. Finally, teams could use a more integrated approach, coding both at the same time and then aligning the qualitative and quantitative findings to see where they agree and where more analysis is needed to explain a mismatch (triangulation).

  – Plans for quantitative data should clearly explain the process for assessing data quality, as well as data cleaning and analysis. It should outline which measures will be used for descriptive data analysis and which for hypothesis testing.

  – Plans for qualitative data should articulate the various methods that will be used for analysis. For example, the plan should explain if or how coding software is to be used, include a draft codebook, and outline any plans for using deductive or inductive coding.

- Beyond describing how quantitative and qualitative implementation data analyses will complement each other, the team should also discuss any plans for having the impact and implementation analyses more noticeably complement or inform one another.

  – If the impact analysis uses individual-level random assignment in multiple sites–each with its own impact estimate–does the implementation analysis provide insights at the site level to identify features likely to be associated with site-level impacts?

  – If the impact analysis uses cluster random assignment (so there are impact estimates only for each cluster), does the implementation analysis provide insights about the clusters (for example, school districts or regions) that might be associated with cluster-level impacts?

- What is your unit of analysis? For example, are you going to be most interested in analyzing networks or collaboratives, organizations (such as schools or centers), individuals (such as students or program participants), or some combination of these? Is your unit of analysis for the implementation component the same as in the impact component (if not, have you acknowledged what the implications might be for what you will be able to say in the report)? And how does the unit of analysis affect both data collection and analysis activities?

Are there sufficient units to estimate the associations of interest among impacts, intervention implementation, context, and contrast?

- Is there variation in both dependent and explanatory measures? Are there plans for running sensitivity tests on findings?

- Will you stage your work so it facilitates learning over time (sequential activities)? For example, can you build in time to analyze data, review what you are learning and what questions remain, and then use the information learned from these activities to inform the next round of data collection?

At an appropriate stage of the project, an implementation researcher outside the team should review analysis plans as a milestone review. This may occur together with the measurement plan, or separately, depending on project needs.

Reporting the findings

This section should describe how the team will report findings. Share findings in ways that will help readers navigate the evidence, place it in context, and use the findings in syntheses to inform future research.

Appendix B. Template for assessing intervention fidelity

Goodson et al. (2019) provides technical assistance for evaluation to recipients of the Institute of Education Sciences' (IES) Education and Innovation Research (EIR) grants. This guidance includes a template for reporting intervention fidelity. Exhibit B.1 provides an excerpt from Goodson et al.'s template.

Exhibit B.1. Excerpt from Goodson et al. (2019) template for reporting intervention fidelity

Indicators	Definition	Unit of implemen-tation	Data source(s)	Data collection (who, when)	Score for levels of implemen-tation	Threshold for adequate implemen-tation at unit level	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to next higher level if needed (score and threshold: Indicate level	Roll-up to sample for fidelity measure (N = # units in which the intervention is being im-plemented)	Expected sample for fidelity measure (N = # units in which the intervention is being im-plemented)	Expected years of fidelity measure-ment
1. Conduct needs assess-ment with principal	Beginning of each school year	Principal	Coaching records kept by grantee	Evaluator, June of each year	1 = yes Assessment conducted, 0 = no					N = 29 principals...	2014-15 = year 2 2015-16 = year 3
2. Leadership coaching provided to principals	Number of days	Principal	Coaching records kept by grantee	Evaluator, June of each year	(Years 2-3) 1 = Principal received 17-22 coaching days, 0 = less than 17					N = 29 principals...	2014-15 = year 2 2015-16 = year 3
3. Key compo-nent score	NA	NA	NA	NA	Sum of indicator scores: 0-2	Principal: Adequate implement-tation = 2	District-level implemented with adequate fidelity = 75% or more principals were at adequate implementation	NA	Adequate = 100% of districts with adequate implemen-tation	All principals in all districts (N = 29 principals in 3 districts)	2014-15 = year 2 2015-16 = year 3

Indicators	Definition	Unit of implemen-tation	Data source(s)	Data collection (who, when)	Score for levels of implemen-tation	Threshold for adequate implemen-tation at unit level	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to next higher level if needed (score and threshold: Indicate level	Roll-up to sample for fidelity measure (N = # units in which the intervention is being im-plemented)	Expected sample for fidelity measure (N = # units in which the intervention is being im-plemented)	Expected years of fidelity measure-ment
					Threshold				100% districts	# of units	Year
					Achieved score at sample level				100% districts	29 principals, 3 districts	2014-15
Fidelity results					Met threshold with fidelity (Yes, No, NA)				Yes		
					Achieved score at sample level				100% districts	29 principals, 3 districts	2016-17
					Met threshold with fidelity (Yes, No, NA)				Yes		

Source: Excerpted from Exhibit 4 in Goodson et al. (2019).

Appendix C. Checking data quality, combining data for analyses, and conducting analyses

As noted in Section V of this guide, quantitative data and analyses can provide useful insight into implementation, context, and contrast. Before analyzing quantitative data, study teams should conduct standard checks for data quality and, if needed, reduce or combine information. Selecting variables for analyses and conducting analyses involves a set of considerations, as discussed below.

Considerations for checking data quality

- **Check for the degree of missing data on individual items and on entire records**. Are any fields or records unusable due to the amount of missing data? Does missingness show patterns? For example, logs might have more missing data during early or final stages of implementation, on certain days of the week, or times of the year, such as near holidays. Ideally, this is something you will be monitoring as you collect data and can work to reduce missingness as a study progresses. However, when analyzing the data, consider whether you have enough useable information for analysis, attempt to seek other sources to fill gaps (for example, reach out to program leaders or educators to see if they have alternative records or sources of data to help fill in the gaps), and when writing about the findings, be sure to document the degree and patterns of missingness, and note whether and how such patterns could plausibly affect interpretations of findings.

- **Use insights from interviews or focus groups to examine quantitative data, especially data collected specifically for the study.** For example, did interviewees comment on the burden or ease of entering data in logs or the intrusive or welcome nature of survey questions? Again, if this is something that you can detect early on, you may be able to make adjustments to data collection protocols and ease the burden or level of intrusiveness. It may also mean not relying on particularly problematic data in your analyses or acknowledging concerns about the data.

Considerations for reducing or combining information

- **Prioritize variables that have more complete data** unless the patterns of missing data are meaningful in some way, or you can reliably impute the missing data or use other techniques to address the missing data.[25]

---

[25] A number of sources provide guidance about imputing missing data and related analyses. Examples include Stuart (2012) and Gelman & Hill, Chapter 25 (2007).

- **Focus on variables that capture key constructs** concisely and do not convey false or unneeded precision. For example, if teachers receive five days of training, reporting the number of training sessions attended is likely sufficient rather than number of minutes of training attended.

- **Identify variables with less measurement error**. For example, when choosing between items about classroom activities, give more weight to an item recorded closer in time to the activity, reducing the likelihood of recall error.

- **Select items for which you can assess reliability** using metrics such as Cronbach's alpha or inter-rater reliability.

- **Choose variables that vary so they can be used to examine associations with impacts.** As candidates for further analyses, focus on variables that are central to the intervention's logic model and exhibit variation. For example, examining variation between tutor background and intervention impacts on students' reading ability would be possible if there is natural variation in the study sample for tutors' educational attainment levels. Such analyses would not be possible if the highest attainment level for all tutors was an undergraduate degree.

- **Consider data reduction strategies such as exploratory or confirmatory factor analyses**. If you aim to reduce the number of individual variables to be analyzed or to identify underlying constructs but are not drawing on a theoretical model or validated measure, consider conducting exploratory factor analysis (Finch, 2020; Watkins, 2018). If you are using a validated measure from previous studies or if you are seeking to test a specific hypothesis about the measures, consider using confirmatory factor analysis (Roos & Bauldry, 2021).

Considerations for analyzing associations

- **Examine associations based on theory, practice, or prior empirical studies.** Hypotheses that draw from prior research and practice knowledge provide stronger justifications than automated searches for empirical associations.

- **Align units of analysis.** Ensure that you estimate associations for the same units of analysis. For example, if the study design results in impact estimates at the school level, then associations of impacts with school-level or higher characteristics would be appropriate to explore. Associations of impacts with teacher- or classroom-level characteristics would not be appropriate, unless aggregated to the school level. It might be appropriate to estimate associations between two measures collected at the teacher level, however, such as years of experience and fidelity to a curriculum model.[26]

---

[26] Multilevel model estimation can maintain the nested structure of the data (for example, teachers clustered within schools) and estimate associations at higher levels of analysis by modeling intercepts or slopes.

- **Ensure you have sufficient units to estimate associations.** A rough rule of thumb for estimating conditional associations in regression analyses (for example, between implementation or context and impacts) is to have 10 observations for each explanatory variable in the model at the appropriate unit of analysis (Raudenbush & Bryk, 2002,). For example, if impacts are estimated for 40 schools, examine at most four explanatory variables. When working with nested data and estimating multilevel associations, however, additional considerations such as intraclass correlations are relevant for model specification and interpretation, and additional caution is needed. Sources with detailed guidance include Dong et al. (2020), Spybrook et al. (2020), Kelcey et al. (2020), and Kelcey et al. (2017).

- **Ensure that both dependent and independent measures vary**. If there is little or no variation in dependent or independent measures, then it is unlikely there will be an association between them. Thus, if either the implementation or impact measures have little variation, it is unlikely the former can explain the latter.

- **Plot the data and run sensitivity tests on findings.** Plotting data can show whether outliers may be driving results. Also, check whether estimates are robust by calculating both unconditional and conditional associations. For example, McCoy et al. (2016) examined whether impacts on children's early language processes and preliteracy outcomes in Head Start centers were associated with urbanicity of communities where the centers were located. They estimate associations between urbanicity and impacts, and also a series of models that explore whether characteristics of the community, Head Start center, or children and families could be confounding their initial findings about associations between urbanicity and impacts.

Appendix D. Reporting guidelines for future syntheses and meta-analyses

This appendix provides examples of reporting guidance developed by federal agencies and research teams.

What Works Clearinghouse (WWC) reporting guide for study authors (IES, 2021b)
https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/WWC_Author_Guide_Jul2021.pdf

This guide lists information that a study should report to be eligible for WWC review. Table 1 lists information about the intervention and comparison conditions, and Table 2 lists information about the study sample and context.

Advancing the use of core components of effective programs: Suggestions for researchers publishing evaluation results (Dymnicki et al., 2020)
https://aspe.hhs.gov/sites/default/files/private/pdf/263931/ASPE-Brief-Core-Components.pdf

This brief suggests priority characteristics to be reported in studies so that research findings can more easily contribute to meta-analyses that seek to identify core components of interventions. Noting that "there is no checklist that is right for every study" (p. 2), the brief suggests priority characteristics of settings, participants, programs, and implementation.

Development of a checklist to assess the quality of reporting of knowledge translation interventions using the Workgroup for Intervention Development and Evaluation Research (WIDER) recommendations (Albrecht et al., 2013)
https://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-8-52

This article describes recommendations from WIDER for reporting details about the content of behavioral interventions to improve health outcomes and reduce health spending. Table 1 lists the recommendations, which may apply to education interventions as well.

Better reporting of interventions: Template for Intervention Description and Replication (TIDieR) checklist and guide (Hoffmann et al., 2014)
https://www.bmj.com/content/348/bmj.g1687

This article recommends what intervention-related information a study should report, with a focus on health research. Table 1 lists the recommended information. The Related Content tab available through the link includes Appendix 3, which is a template with the recommended fields, and Appendix 4, which includes examples of different reporting formats.

Developing standards for reporting implementation studies of complex interventions (StaRI): A systematic review and e-Delphi (Pinnock et al., 2015)
https://implementationscience.biomedcentral.com/articles/10.1186/s13012-015-0235-z

> *This article describes reporting standards for implementation studies (not necessarily ones that are conducted in the context of a randomized controlled trial) in health settings. Table 2 lists information on how an expert panel reached consensus.*

Contextual factors: The importance of considering and reporting on context in research on the patient-centered medical home (Stange & Glasgow, 2013)
https://www.ahrq.gov/sites/default/files/wysiwyg/ncepcr/tools/PCMH/contextual-factors.pdf

> *As its title suggests, this guide for researchers focuses on gathering information about context and reporting about it in studies that involve patient-centered medical homes.*

Acknowledgements