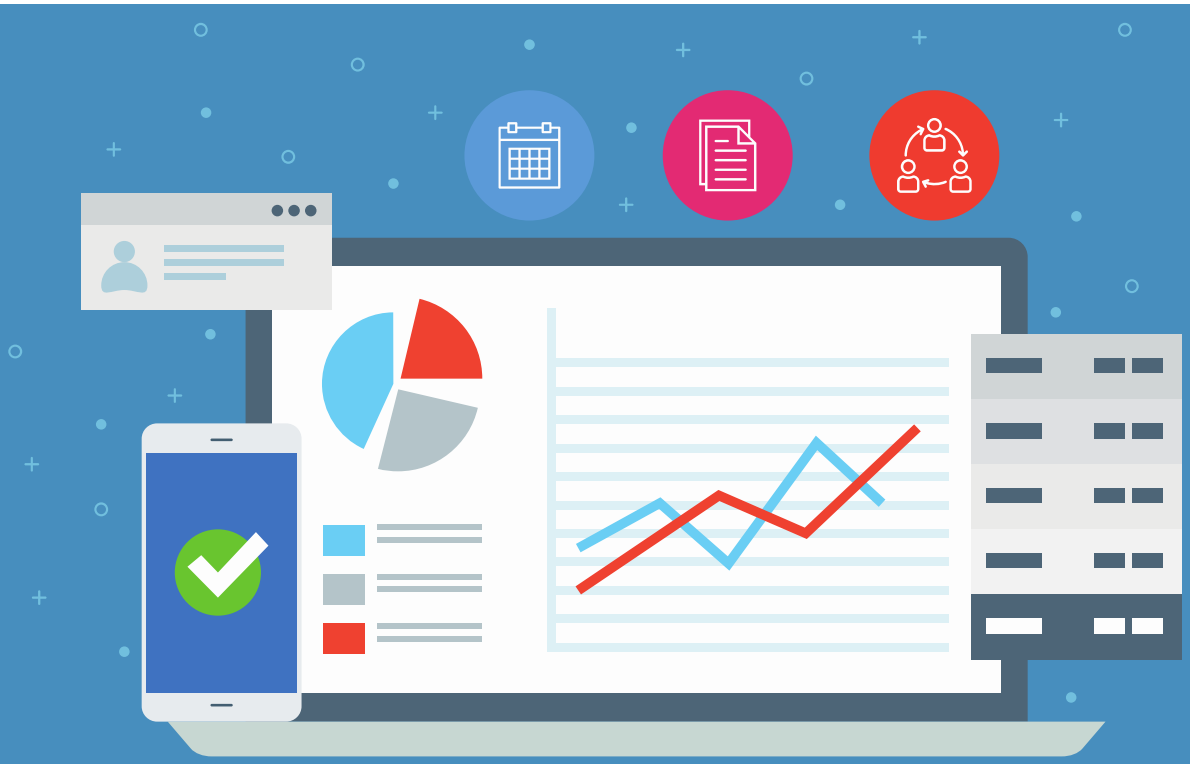


TOOLKIT | December 2018

Formative Evaluation Toolkit

A Step-by-Step Guide and Resources for Evaluating Program Implementation and Early Outcomes

TOOLKIT | December 2018

Formative Evaluation Toolkit

A Step-by-Step Guide and Resources for Evaluating Program Implementation and Early Outcomes

Submitted to

Chereese Phillips, Ph.D.
Child Welfare Program Specialist
Children's Bureau
Administration for Children and Families
U.S. Department of Health and Human Services

Prepared by

James Bell Associates
3033 Wilson Blvd., Suite 650
Arlington, VA 22201
(703) 528-3230
www.jbassoc.com

Elliott Graham, Ph.D.
Project Director

Disclaimer

This publication was developed by James Bell Associates on behalf of the Children's Bureau, Administration for Children and Families (ACF), U.S. Department of Health and Human Services (HHS), under contract number HHSP233201500133I, order number HHSP23337001T. Its contents are the sole responsibility of the author and do not necessarily represent the official views of the Children's Bureau, ACF, or HHS.

For more information, please visit the Children's Bureau website at https://www.acf.hhs.gov/cb.

Children's Bureau

jba JAMES BELL ASSOCIATES

Acknowledgments

The authors wish to thank several individuals for their support in the production of this toolkit. Catherine Heath of the Children's Bureau, U.S. Department of Health and Human Services, was instrumental in spearheading this effort and helped inform its development through her work with Children's Bureau discretionary grantees that received Planning Grants to Develop a Model Intervention for Youth and Young Adults with Child Welfare Involvement At-Risk of Homelessness (HHS-2013-ACF-ACYF-CA-0636). We are also grateful to Serena L. Williams, Ph.D., Matthew McGuire, and Melinda Baldwin, Ph.D., of the Children's Bureau; to Jan P. Shafer, Director of the Children's Bureau's Division of Program Innovation; as well as to John Denniston, formerly with ICF/Child Welfare Information Gateway, for their insights and guidance. We would also like to express our appreciation to the technical assistance and evaluation teams of the Permanency Innovations Initiative; their work was influential in developing the content of many sections of the toolkit.

Contents

Exhibits

Introduction

Few evaluations of social programs have demonstrated positive results (Laura and John Arnold Foundation, 2017). Part of the problem may be that program teams and evaluators often conduct rigorous summative evaluations of programs that are not actually ready for that level of scrutiny (Epstein & Klerman, 2013).

This toolkit introduces formative evaluation, a method for evaluating programs during early implementation to inform program improvement and assess readiness for rigorous summative evaluation. The toolkit is designed primarily for child welfare agencies and professionals to use in partnership with program evaluators, but others in human services may also find it useful. The content aligns with the experiences of state, local, and nonprofit child-serving organizations and evaluators funded through the Children's Bureau discretionary grant program.

The toolkit guides users through the process of formative evaluation. It has three sections:

I. **Understanding Formative Evaluation** provides an overview of formative evaluation, explains how it differs from summative evaluation, and describes the role of formative evaluation in the evidence-building process.
II. **Meeting Preconditions of Formative Evaluation** poses questions and provides resources to help programs lay the groundwork for formative evaluation.
III. **Conducting Formative Evaluation** outlines the steps to design and conduct formative evaluation and use the findings to improve programs and determine readiness for summative evaluation.

The toolkit includes—

- Examples based on a fictional agency, the Washington County Department of Human Services, at the end of each section[1]
- Tools you can complete with your own program's information using the Washington County examples as a guide (appendix A)
- Additional resources cited throughout the toolkit
- A glossary of key terms (appendix B)

[1] Washington County is used throughout this toolkit for illustrative and educational purposes only. Any similarities with agencies located in actual Washington Counties in the United States, or with human service agencies in any other U.S. counties, are entirely coincidental.

I. Understanding Formative Evaluation

Formative evaluation is typically conducted on an early version of a program, while the program is still being developed and improved (Scriven, 1997). Summative evaluation is conducted later, once the program has reached full implementation (exhibit 1).

The purpose of formative evaluation is to provide feedback to program directors and staff about program functioning and some short-term outcomes (also known as proximal outcomes). Formative evaluation can determine whether the program is being implemented as intended and producing expected outputs, and it can reveal whether short-term outcomes are trending in the right direction. The information helps program staff improve and refine the program early and keep it on track.

One way to think about the difference between formative and summative evaluation is to think of a chef preparing a meal. Any good chef will taste the food as he or she prepares it, and make adjustments as needed. Once the meal is served, it is out of the chef's hands. Similarly, formative evaluation is a time to check and adjust program functioning before moving on to summative evaluation.

Exhibit 1. Key Differences Between Formative and Summative Evaluation

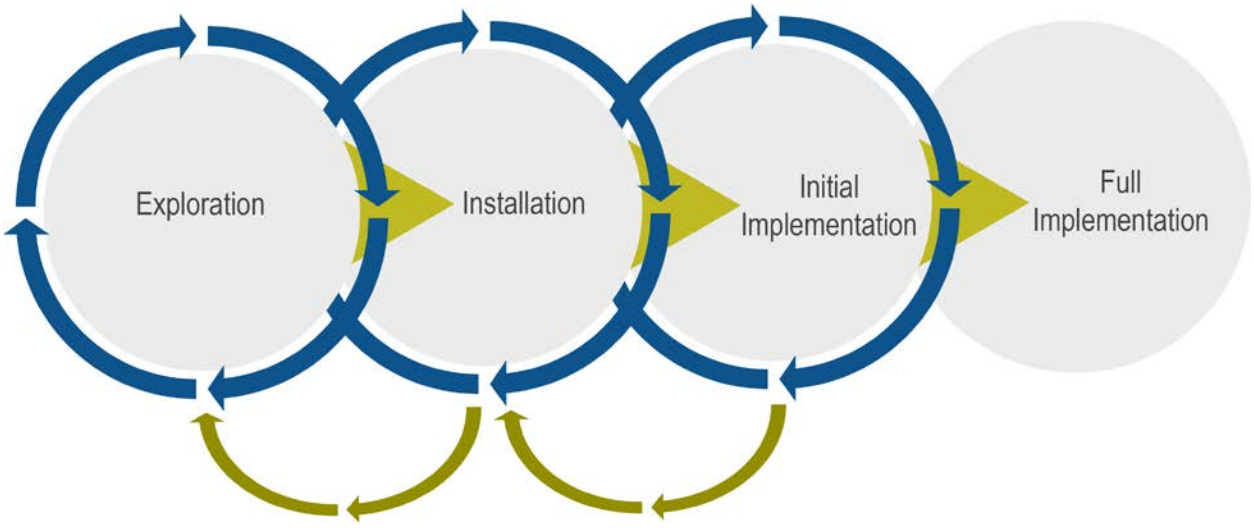
Question	Formative evaluation	Summative evaluation
What is it?	An evaluation of whether a program is producing expected outputs and short-term outcomes; it may also provide information about unintended outcomes	A rigorous evaluation of short-, medium-, and long-term outcomes
When is it conducted?	During initial implementation, while a program is still being developed	During full implementation, once a program is fully functioning and no additional changes are expected
Who conducts it?	Program staff or an external evaluator	Usually an external evaluator

Question	Formative evaluation	Summative evaluation
What questions can it answer?	Are key aspects of the program functioning as intended? Are early outcomes moving in the right direction?	Did the program have the intended effect on its target population? Are the outcomes attributable to the program?
What outputs and outcomes are measured?	Key outputs and short-term outcomes only	All outputs and outcomes
How often is it conducted?	Often repeatedly	Usually just once
Who are the results shared with?	Primarily program staff	Program staff, funders, key stakeholders, and possibly a wider audience
How many cases should it include?	Small group of cases is acceptable; tests of statistical significance are usually not necessary	Power analysis should be used to determine the sample size needed to test for statistical significance

Note: Online resources for statistical power analyses include Sample Size Calculators from the University of California, San Francisco, Clinical and Translational Science Institute.

Program implementation is a process that unfolds over time—in child welfare, it may take several years. In this toolkit, we refer to four overlapping stages of implementation (exhibit 2): (1) exploration, (2) installation, (3) initial implementation, and (4) full implementation (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005; Permanency Innovations Initiative Training and Technical Assistance Project, 2013). Formative evaluation takes place during the third stage, initial implementation. Activities within each stage are described below.

Exhibit 2. Stages of Implementation



1. Exploration

Exploration may be the most important stage because it sets a program on its trajectory. It is easier to get things right during exploration than to get back on track later.

The exploration stage lays the groundwork for a new program. Staff determine needs and goals, assess agency readiness and resources to take on a new program, identify a target population, articulate a theory of change, and explore the fit of potential programs. An implementation team and key stakeholders are brought to the table. Once a program is selected, the team develops a logic model and identifies potential implementation barriers and solutions.

Exploration usually culminates in the development of a written implementation plan and initial ideas about evaluation.

Exploration may be the most important stage because it sets a program on its trajectory. It is easier to get things right during exploration than to get back on track later.

2. Installation

Installation focuses on refining teaming structures and processes and operationalizing the program, including aspects such as staff hiring and training, communication, and administration. The trajectory of program implementation depends largely on whether you are implementing a manualized program or an innovation (see box).

Evidence-based programs are often manualized; the developer provides a program manual or practice profile and other materials to support implementation. In contrast, untested innovations do not typically have ready-made materials. You must develop the materials yourself before proceeding to initial implementation and formative evaluation. It may take more time than anticipated before an innovation is ready to move into the next stage, initial implementation, when formative evaluation happens.

Evidence-Based Programs Versus Innovations

Evidence-based programs (also known as evidence-supported interventions) have demonstrated effects on children and families or are based on research. Evidence is built over time through rigorous and repeated data collection and analysis. As program components become better understood and refined, a program may eventually be manualized so it can be replicated with fidelity (Fraser, Richman, Galinsky, & Day, 2009). Manualized programs typically provide detailed instructions and materials about components that affect children and families directly (for example, the number of treatment hours) and indirectly (for example, procedures for hiring and training staff, collecting data, and assessing fidelity).

Innovations are programs that have not been tested or have been adapted from other contexts. Because relatively few evidence-based programs are designed for use in a child welfare context, many child welfare programs are innovations (Testa & White, 2014). Formative evaluation tells you whether an innovation is on the right track.

3. Initial Implementation

Initial implementation begins when the first clients receive program services. Early in this stage, the implementation team conducts usability testing, also known as rapid-cycle improvement processes. Through careful observation, data collection, and interpretation, staff repeatedly assess program functioning and make quick improvements. The program may then be ready for **formative evaluation**, which often reveals additional areas for improvement before full implementation begins.

4. Full Implementation

Full implementation begins once program services, structures, and processes are stable. Staff continue to monitor implementation and respond to challenges.

If a formative evaluation indicated program outputs and short-term outcomes were trending in the right direction, the program may be ready for **summative evaluation**. Once the decision is made to begin summative evaluation, no more modifications to the program model should be made. The services and service delivery approach used with the last child or family enrolled should mirror those used with the first child or family enrolled.

Before proceeding to section II, see the overview below of the fictional program that serves as the basis for the examples in this toolkit. The Washington County Department of Human Services is implementing an in-home parenting program aimed at improving parents' capacity to safely care for their children and reducing removals and maltreatment reports.

Washington County In-Home Parenting Program

Washington County was concerned about its relatively high rate of families with multiple reports of child neglect. It convened a team to explore the needs of children and families in the county; identify the population most at risk (e.g., low-income, younger parents with young children); and gather staff and stakeholder input on promising strategies and possible barriers. The team then researched programs that had strong evidence of reducing risk for neglect and were a good fit for the agency and target population.

The team decided to implement an adaptation of SafeCare, an evidence-based home visiting program. It is an intensive, 8-week program for parents with children aged 0–10 at risk of removal. Staff conduct a family assessment at intake to determine referral to the program. Family therapists certified in SafeCare contact families within 24 hours of referral and schedule in-home visits twice a week for up to 8 weeks. Each visit lasts 1–2 hours and includes a variety of service strategies, including cognitive-behavioral therapy; parenting skills training (e.g., active listening, modeling positive parenting, coaching parents in new skills); development of an in-home safety plan; and referral to additional community-based services. Parents are encouraged to attend weekly parent support groups hosted by the program. They complete additional risk assessments at program completion and 6 and 12 months after completion.

The goal of the program is to improve parents' capacity to safely care for their children so fewer children are removed and fewer re-reports or new reports of maltreatment are made.

II. Meeting Preconditions of Formative Evaluation

Formative evaluation requires essential groundwork during the first three implementation stages introduced in section I: **exploration, installation,** and early **initial implementation**. The questions in section II will help you determine whether you have completed these key activities and are ready for formative evaluation.

The following tools are referenced in this section and included in appendix A. Examples based on the fictional Washington County program are included at the end of this section.

II.1. Readiness for Formative Evaluation Checklist

II.2. Refining Your Theory of Change Tool

II.3. Refining Your Logic Model Tool

II.4. Usability Testing Tool

As you consider each precondition below, see the resources listed in the Readiness for Formative Evaluation Checklist (tool II.1) to learn more.

1. Do you have an implementation team in place?

The implementation team oversees the implementation and formative evaluation of the program. It researches and selects the program, drafts workplans, analyzes data, and sustains the program in practice. The team documents its work and decisions in an implementation plan and evaluation plan.

You may have subteams or shifting team membership, depending on the stage of implementation. The important thing is to include the right people—those who are involved in the implementation and evaluation and can identify and address potential challenges.

Key Implementation Team Members

- Agency leaders and decision makers
- Program staff and supervisors
- Staff with knowledge of program data and continuous quality improvement
- Evaluators
- External stakeholders

2. Have you set realistic time lines?

The steps leading to formative evaluation typically take a long time to complete. In child welfare, it may be 2 or more years before a program is ready for formative evaluation. When developing workplans, consider the type of program (manualized versus innovation; see section I) and the time needed to meet the preconditions described in this section. Also see the Washington County Formative Evaluation Time Line example.

The steps leading to formative evaluation typically take a long time to complete. In child welfare, it may be 2 or more years before a program is ready for formative evaluation.

3. Do you have a clear theory of change based on data?

It is critical to dig into your organization's data to articulate a clear theory about the root cause of the identified problem, the desired outcomes, and a way to reach those outcomes. Without fully understanding the problem, the affected target population, and barriers to addressing the problem, you risk selecting a program that will be ineffective. Situational or root cause analysis—that is, posing and testing hypotheses by mining existing data or gathering new data—provides this knowledge. See the resources on root cause analysis in tool II.1.

A theory of change uses the results of a root cause analysis to identify causal linkages—that is, a program or activities that will change outcomes, and the assumptions about how and why the changes will occur (Funnell & Rogers, 2011). The theory of change guides the selection of a

program to drive the needed changes and informs the logic model (see #4 below). See the resources on theory of change in tool II.1.

Theories of change are challenging to articulate and can change as new information is gathered throughout program implementation (Dhillon & Vaca, 2018). Revisit and critically evaluate your theory of change often, including before formative evaluation. Use the Refining Your Theory of Change Tool (II.2) and see the Washington County Theory of Change example.

4. Have you selected a program that fits your needs and aligns with your theory of change?

The implementation team typically reviews an array of possible programs before selecting the best one to achieve the desired outcomes. See the resources on program assessment and selection in tool II.1.

Does the program you selected—

- Align with the theory of change?
- Have evidence that it—
  - Achieves the desired outcomes?
  - Addresses the needs of the target population?
  - Has been replicated?
- Have a manual outlining core components for implementation or adaptation?
- Fit with the agency and system and have staff buy-in?

If the program selected was designed and tested for a different population or context, it may require adaptation for use in child welfare. Adaptation is the process of modifying a program to meet local characteristics without changing its core components. When adapting a program, it is critical to consult the program developer and conduct usability testing (see #8 below).

5. Have you developed a logic model and evaluated it for plausibility and gaps?

Once you have selected a program, you can build your logic model. A logic model—

- Visually depicts how the program should operate
- Expresses goals from the theory of change in terms of outputs and outcomes
- Outlines the data collection strategy

The logic model should ensure evaluability—that the program is defined well enough to be implemented with fidelity, has plausible objectives, and is ready and suitable for rigorous evaluation. Setting benchmarks for outputs and short-term outcomes will tell you whether the program is working as intended. Benchmarks are typically set before a formative evaluation begins and are examined during early usability testing (see #8 below) and formative evaluation.

Review the logic model to identify gaps in logic, faulty assumptions, and unrealistic expectations. A high-quality logic model built on a strong theory of change may be the most important product of the exploration stage, because it strengthens the likelihood of achieving expected outcomes. Undetected flaws in a program's logic can cause major problems later.

Use the Refining Your Logic Model Tool (II.3) to improve the plausibility and utility of your logic model, and see the Washington County Logic Model example. For more information, see *Developing a Logic Model* (James Bell Associates, 2007) and *Guide to Data-Driven Decision Making* (James Bell Associates, 2018).

6. Have you operationalized the program and installed supports?

A program manual or practice profile helps staff operationalize a program. It explains what to do, how to do it, and how to determine whether you are doing it correctly. It also helps evaluators identify core elements that should be measured to assess fidelity and factors that contribute to or undermine effectiveness.

If you selected an evidence-based program, a manual may already be available. If you are adapting a program or developing an innovation, you may need to modify or create a manual. See the resources on operationalizing programs in tool II.1.

Does your program manual or practice profile include—

- **Essential functions** or activities that set the program apart from usual practice and are backed by evidence?
- **A definition for each function** that aligns with the theory of change and goals?
- **Behaviorally based practice indicators**—that is, observable and measurable indicators that staff are performing activities correctly?
- **Practice criteria** that describe levels of performance on a continuum that can be applied in coaching and supervision to improve staff performance?

Key implementation supports to install or strengthen during the installation stage include—

- Leadership and stakeholder engagement
- Communication of goals
- Staff recruitment and retention
- Staff training
- Staff coaching, modeling, and feedback
- Fidelity assessment tools and protocols
- Data systems (see #7)

Measuring implementation supports is essential to the evaluation strategy. Frequently, an evaluation will indicate that a program was ineffective, but further examination shows the organization failed to fully implement the program, in part because of inadequate supports. The implementation team should assess and test implementation supports (see #8 below) and incorporate ongoing measurement of supports in the formative evaluation.

See the resources on implementation supports in tool II.1. For additional information, see the *Guide To Developing, Implementing, and Assessing an Innovation* (Permanency Innovations Initiative Training and Technical Assistance Project, 2016) and visit the National Implementation Research Network website.

7. Have you identified needed data elements and installed data systems?

The implementation team needs data throughout the four stages of implementation to—

- Ask and answer important questions related to the target population, risk characteristics, and desired outcomes
- Monitor implementation and inform improvements
- Test and document outcomes in formative and summative evaluation

Plan and begin to install a decision support data system during early implementation. See the *Guide to Data-Driven Decision Making* (James Bell Associates, 2018), *Data System Improvement Toolkit* (Geary, Poes, Iannone-Walker, Porter, Callis, Buckless, & Day, 2018), and additional resources on data systems.

8. Have you conducted usability testing to address "bugs"?

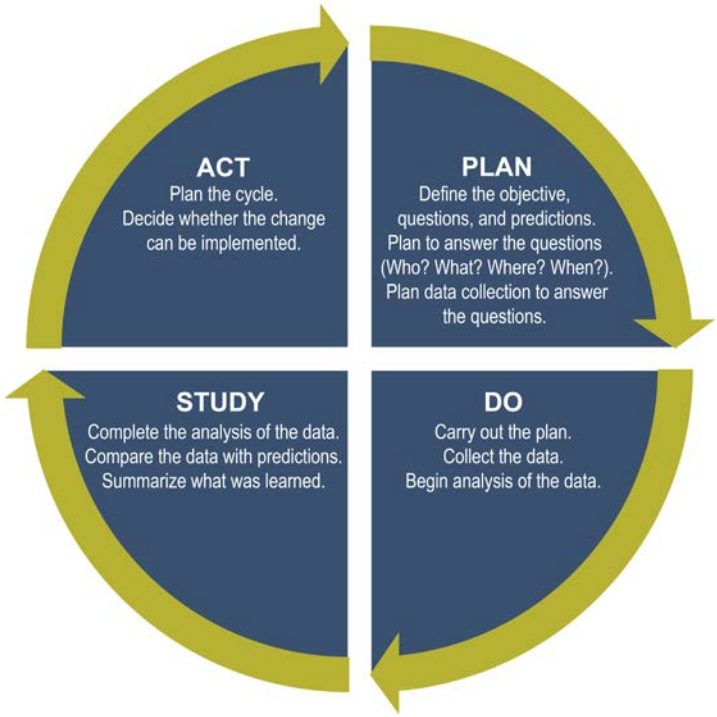
During the initial implementation stage, the implementation team may conduct usability testing or rapid-cycle improvement. Testing program components under real-world conditions reveals whether they function as planned and identifies issues. For example, you can test your program enrollment process, child assessment process, or new data system. Usability testing helps get tricky program components right before moving to full implementation and formative evaluation.

Usability testing is particularly important when you—

- Are implementing an innovation, adaptation, or new program component
- Are implementing a program in coordination with partners
- Set challenging time lines for program installation

Usability testing shares concepts with data-driven decision making and continuous quality improvement, which incorporate the Plan-Do-Study-Act framework (exhibit 3).[2]

Exhibit 3. Plan-Do-Study-Act Framework



[2] See Deming (1986) or PDSA Worksheet (Institute for Healthcare Improvement, 2018) for an in-depth description of the framework.

By testing only one or a few components at a time, you obtain efficient feedback to inform early modifications. The team defines questions and targets for success (exhibit 4) and then designs and conducts short-term rounds of "mini-evaluations," making improvements between cycles.

Exhibit 4. Usability Testing: Sample Questions and Targets

Area	Questions	Targets
Referral and enrollment	Are referral criteria well operationalized and understood by those who will refer families?	Percentage of referred families that met eligibility criteria (80%)
Core components	Is there evidence staff are performing the program's essential functions?	Percentage of staff who scheduled appointments with participating families within 7 days of enrollment (80%)
Implementation	Do staff conduct baseline assessments on time?	Percentage of baseline assessments completed within 14 days of enrollment (80%)
Data collection	Are protocols for data collection processes feasible? Are data entry and reporting occurring as intended?	Percentage of assessments correctly entered into data system within 7 days (80%)

Usability cycles often only last from a few weeks to a few months. Data must be collected and interpreted quickly. Continue the cycle of feedback and improvement until you meet your targets. The program is then likely ready for **formative evaluation**.

Use the Usability Testing Tool (II.4) to make decisions about testing and see the Washington County Usability Testing Plan and Results example. For additional information, see the National Implementation Research Network website and *Guide to Developing, Implementing, and Assessing an Innovation* (Permanency Innovations Initiative Training and Technical Assistance Project, 2016).

Tips

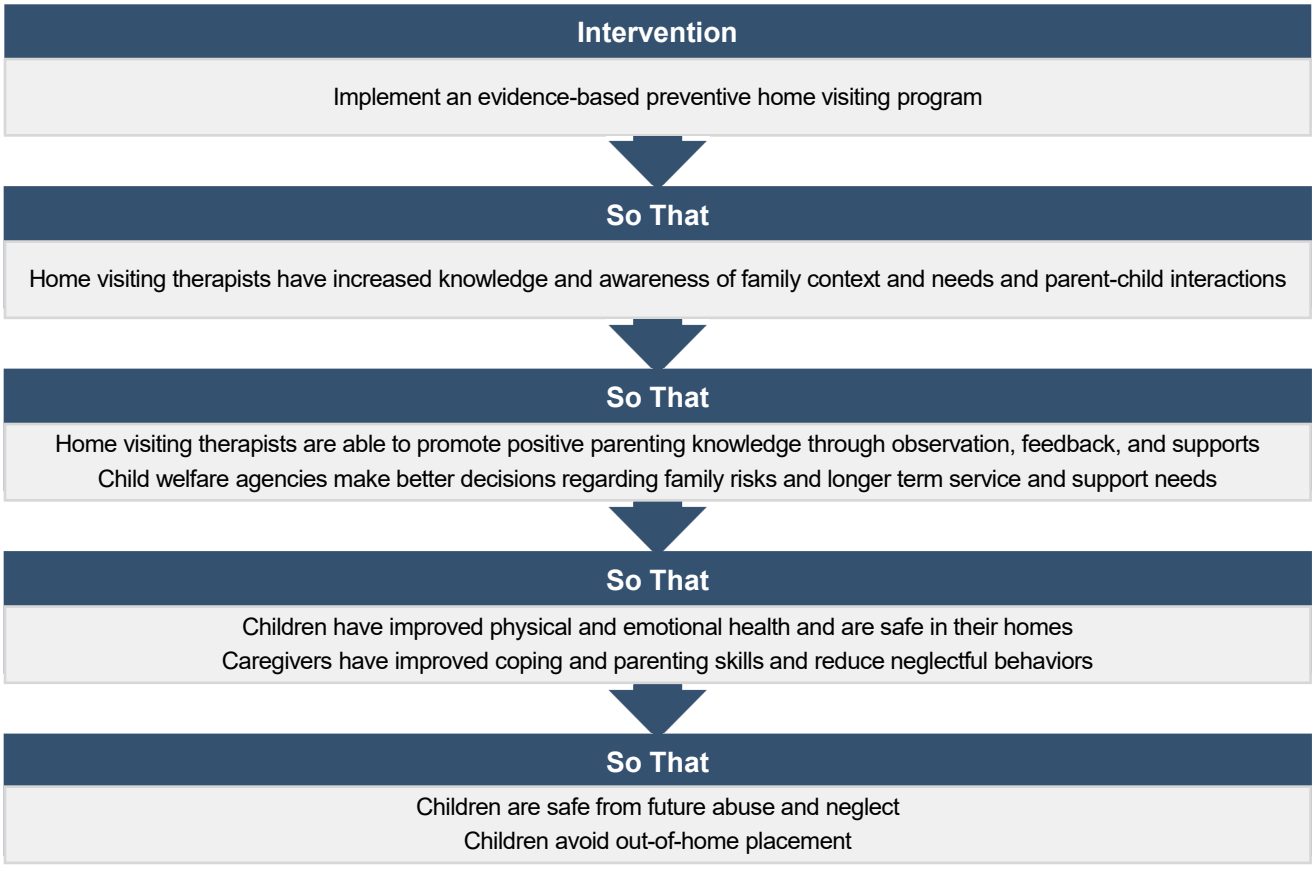
- Test processes that are challenging, new, or require coordination with partners.
- Engage in just enough preparation to get started—tests don't have to be perfect.

Source: State Implementation and Scaling-Up of Evidence-Based Practices Center and National Implementation Research Network, 2017

Section II Examples

Washington County Theory of Change

The Washington County implementation team began developing its theory of change by specifying the causal links between program activities and desired outcomes in an outcomes chain, presented below. See appendix A for a tool you can use to refine your own theory of change.



The links in the chain are connected through a series of "so that" statements to articulate temporal and logical relationships between actions and outcomes.

Once the chain was formed, the implementation team articulated a narrative theory of change:

By implementing a proven preventive intervention that occurs in families' homes, child welfare professionals will better understand families' needs and contextual factors, and promote positive parent-child interactions and increased family supports through individualized coaching and modeling strategies that fit the families' environments. As a result, caregivers will have improved coping and parenting skills, and their children will experience better behavioral, health, and well-being outcomes, including reduced risk for neglect, abuse, and out-of-home placement.

Washington County Logic Model

See appendix A for a tool you can use to refine your own logic model.

Washington County Usability Testing Plan

Washington County used the table below to plan usability testing. See appendix A for a blank version you can use to plan your own usability testing.

Usability test(s)	Area	Usability question	Scope of testing	Participants	Roles	Metrics or key outputs	Data sources and methods	Time line	Criteria for completion
	Which area are you targeting?	*What are you trying to learn?*	*Which process or step is being tested?*	*Who will be included in the test and how will they be identified or included?*	*Who are the—* • *Decision makers?* • *Test coordinators?* • *Data collectors?* • *Analysts and reporters?*	*What are the outputs that could be assessed in the short term to answer the questions?*	*What data will be collected to measure the metric? What methods will be utilized?*	*When will the test begin and end?*	*What are the criteria for success? What is the benchmark?*
#1	Recruitment, referral, and enrollment processes	Are the recruitment criteria well operationalized and understood by those who will refer?	Eligibility criteria protocols used by all partner agencies	Partner and agency intake staff (n = 6) who are designated project referral sources	Implementation Team will decide on testing success DHS quality assurance team will analyze data	Families referred to the home visiting intervention meet eligibility criteria	Intake referral checklist; family enrollment data forms	3-month period from start date	80% meet eligibility
#2	Core components	Is there evidence that staff are performing essential functions (core components) of the intervention?	Core component of early engagement and relationship building	In-home therapists assigned to cases		Staff schedule home visits with participating families within 7 days of enrollment	Case management database; dates of contact for initial home visit		80% of staff schedule initial visits within identified time period

Usability test(s)	Area	Usability question	Scope of testing	Participants	Roles	Metrics or key outputs	Data sources and methods	Time line	Criteria for completion
	Which area are you targeting?	*What are you trying to learn?*	*Which process or step is being tested?*	*Who will be included in the test and how will they be identified or included?*	*Who are the—* • *Decision makers?* • *Test coordinators?* • *Data collectors?* • *Analysts and reporters?*	*What are the outputs that could be assessed in the short term to answer the questions?*	*What data will be collected to measure the metric? What methods will be utilized?*	*When will the test begin and end?*	*What are the criteria for success? What is the benchmark?*
#3	Implementation	Do client baseline assessments occur on time?	Assessment data used in data-driven decision making for case plans	Program intake staff and in-home therapists responsible for assessments	Implementation Team will decide on testing success DHS quality assurance team will analyze data	Home visit baseline assessments completed within 14 days of enrollment	Case management database; dates of assessment administration	3-month period from start date	80% of baseline assessments completed within identified time period
#4	Data collection	Are protocols for data collection processes feasible? Is data entry and reporting occurring as intended?	Family data available for evaluation and program decision making	Program and evaluation staff responsible for entering data		Assessments correctly entered into data system within 7 days	Case management database; dates of assessment data entry		80% of families with assessments entered within identified time period

Washington County Usability Testing Results

Usability test(s)	Area	Usability metric and benchmark	Results	Repeat testing?	Actions	Roles
	Which area are you targeting?	*What are you trying to achieve?*	*What were the results?*	*Will another round of testing be needed?* (Yes/No)	*What action will be taken in light of these results?*	*Who is responsible for taking the identified actions?*
#1	Recruitment, referral, and enrollment processes	Families referred to the home visiting intervention meet eligibility criteria (80%)	• 80% of families met at least five of the eight criteria; however, some confusion regarding the definition of the criteria. • Results confirmed some data needed for eligibility determination are difficult to access during intake period	Yes	Clarification of eligibility criteria; more in-depth training on eligibility determination; identification of alternate sources of information for those data elements that are difficult to confirm early in the case	Program director
#2	Core components	Staff schedule home visits with participating families within 7 days of enrollment (80%)	82% of families had a home visit within 7 days; however, only 68% completed visit at initially scheduled time	Yes	Identification of additional engagement methods to help ensure families followed through on home visits (motivational interviewing used in early engagement conversations, offering more flexibility to schedule evenings and weekends)	Program director and in-home therapists

Usability test(s)	Area	Usability metric and benchmark	Results	Repeat testing?	Actions	Roles
	Which area are you targeting?	*What are you trying to achieve?*	*What were the results?*	*Will another round of testing be needed?* (Yes/No)	*What action will be taken in light of these results?*	*Who is responsible for taking the identified actions?*
#3	Implementation	Home visiting baseline assessments completed within 14 days of enrollment (80%)	100% of families with a completed home visit completed assessments; however, only 75% of scheduled visits completed within the 14-day period	No	Successful completion of assessments among those families that actually engaged in the home visit suggests the main challenge is getting families to complete the visit rather than the assessment. • Metric will continue to be monitored during the formative evaluation	Quality assurance team to monitor during evaluation
#4	Data collection	Assessments correctly entered into data system within 7 days (80%)	100% of families with completed home visits and completed assessments had data entered on time into the data system	No	No action steps needed	Quality assurance team to monitor during evaluation

III. Conducting Formative Evaluation

If your program has reached the initial implementation stage and met the preconditions for formative evaluation (see section II), you may be ready to conduct a formative evaluation. Recall that the purpose of formative evaluation is to identify problems with program functioning, determine whether outcomes are trending in the desired direction, and assess readiness for summative evaluation.

This section will guide you through the steps of formative evaluation: (1) identifying key formative evaluation questions, (2) selecting indicators of program functioning, (3) choosing evaluation methods, (4) collecting data, (5) analyzing and reporting data, and (6) deciding what comes next.

Remember—formative evaluation is the last opportunity to modify a program before full implementation and summative evaluation. See the Washington County Formative Evaluation Time Line example.

The following tools are referenced in this section and included in appendix A. Examples based on the fictional Washington County program are included at the end of this section.

III.1 Program Enrollment Tracking Tool

III.2 Response Rate Tracking Tool

III.3 Short-Term Outcome Tracking Template

III.4 Formative Evaluation Questions and Indicators/Benchmarks

1. Identify Key Formative Evaluation Questions

The first step in formative evaluation is to decide on your evaluation questions. Consider four key questions.

Is the program reaching the intended number of participants?

It is important to know early whether your program is meeting enrollment and retention goals. In the initial implementation stage, members of your target population begin enrolling in the program and are then tracked. Your participant data may reveal challenges with, for example, your referral process or your ability to retain participants. See the Program Enrollment Tracking Tool (III.1).

How are inputs contributing to program functioning?

Inputs are the resources needed to implement your program, including personnel, materials, space, time, and organizational supports. They should be listed in the resources or inputs column of your logic model (see the Refining Your Logic Model Tool, II.3). Common input categories and indicators are listed in exhibit 5.

Exhibit 5. Examining Common Input Categories

Common input categories	Example indicators
Personnel	• Sufficient staff to conduct program activities • Staff training • Supervisor coaching and oversight
Materials	• Sufficient materials for participants • Culturally competent materials (e.g., language translations) • Accessible materials (e.g., audio descriptions or braille materials) • List of materials adapted or awaiting adaptation
Services	• Appropriate pace of enrollment in services • Availability of services when needed • Accessibility of services (e.g., transportation options)
Time and space	• Sufficient staff time to conduct program activities • Sufficient supervisor time to manage program activities • Appropriate space for program activities
Organizational supports	• Effective scheduling and management of staff to ensure coverage of program activities • Leadership monitoring of program activities • Appropriate data systems

Is the program being delivered as intended?

Formative evaluation can help you determine early if your program is being delivered as intended—that is, with fidelity to the model. In the early implementation stage, it is normal and expected that practitioners may not yet be delivering the program with high fidelity. Through training and coaching, fidelity should increase over time. Use your formative evaluation to examine five key dimensions of fidelity, as detailed in exhibit 6. To learn more, see *Measuring Implementation Fidelity* (James Bell Associates, October 2009).

Exhibit 6. Dimensions of Implementation Fidelity

Fidelity dimension	Description	Common indicators	Recommendations
Adherence	The extent to which core components are delivered as prescribed by the model	• Program content • Methods • Activities	Since all core components should be represented as activities in your logic model, review the research questions developed for the outputs in your logic model.
Exposure	The amount of program delivered in relation to the amount prescribed by the program model; also referred to as dosage	• Number of sessions • Attendance • Frequency and duration of sessions	Review your enrollment tracking table and add information about individual participants. Also review the research questions generated for the outputs in your logic model.
Quality of delivery	The expertise with which providers deliver content	• Provider preparedness • Use of relevant examples • Enthusiasm • Interaction style	Conduct observation using methods such as a structured tool and participant feedback surveys.
Participant responsiveness	The way participants react to or engage in a program (participant satisfaction is one example of responsiveness)	• Participant level of interest in the program • Perceptions about the relevance and usefulness of the program	Use tools such as pre-post surveys that ask about participant experiences with the program.
Program differentiation	The degree to which the critical components of the program are distinguishable from each other and from other programs	• Staff training and program manuals on critical program components and eligibility requirements • Types of program firewalls (e.g., training at the staff, unit, or office level)	Ensure staff understand how participants in the program differ demographically from enrollees in other programs so ineligible cases do not receive the intervention. If using a comparison group, monitor whether only the intended participants receive services.

Are short-term outcomes promising?

Formative evaluation is a chance to examine how key outcome indicators have increased or decreased as expected—for example, from baseline to follow-up. It is also an opportunity to determine whether the program is creating any unintended consequences or harmful outcomes for participants. See the Short-Term Outcome Tracking Template (tool III.3).

2. Select Indicators of Program Functioning

Guided by the evaluation questions, the next step is to select elements of your logic model that will be the focus of your evaluation and tell you how your program is functioning. These might include indicators of implementation activities, program outputs, key short-term outcomes, and possibly some intermediate outcomes. For each indicator, determine a benchmark for success—an expected standard to which you will compare your program's actual functioning.

For example, in the fictional Washington County example featured in this toolkit, the implementation team chose to test all the activities, outputs, and short-term outcomes in its logic model (see Washington County Formative Evaluation Questions and Indicators/Benchmarks).[3] The team set benchmarks based on experience, staff capacity, and best practices; see examples in exhibit 7.

Exhibit 7. Washington County Example Indicators and Benchmarks

Indicators of program functioning	Formative evaluation questions	Benchmarks
Proportion of intake family assessments and ongoing risk assessments completed as scheduled	How many intake family assessments are completed each week? Are they completed on time?	• 85% of families with a screened-in maltreatment case will complete a family assessment within 24 hours of case opening
	How many ongoing risk assessments are completed with participating parents? Are they completed on time?	• 90% of enrolled families will complete a risk assessment at program completion • 85% of enrolled families will complete a risk assessment at 6 months post-program • 75% of enrolled families will complete a risk assessment at 12 months post-program

[3] Your formative evaluation does not necessarily need to include all outputs and short-term outcomes identified in the logic model. Consider which ones are most important to assess in the formative evaluation.

Selecting indicators and setting benchmarks help you think about what program success looks like. For example, ideally, 100 percent of families in Washington County with a screened-in maltreatment case would be assessed within 24 hours, but the implementation team knew this would not be realistic. Instead, it set 85 percent as a reasonable and acceptable benchmark.

If you are implementing a manualized program, the developers should have provided guidelines and parameters for each program component. Review the program materials or contact the developers for information to help you set your benchmarks. If you are implementing an innovative program or an adaptation of an intervention from another field, you will need to work with your implementation team to set benchmarks. Consider your target population, staff capacity, best practices from the literature, and lessons learned from other programs and your practice.

3. Choose Evaluation Methods

Once you have selected indicators of program functioning, determine the methods you will use to study them. This includes determining how you will collect data on those indicators, for how long, and what comparisons you will make to understand whether and how your program participants have been affected.

Formative evaluation should be structured in a way that provides you with useful information relatively quickly. First, decide how many cases you want to include in your sample. The sample size needed for formative evaluation is typically smaller than what may be needed for a summative evaluation (i.e., dozens of cases, not hundreds). Second, choose methods that are feasible to implement in a limited time period (i.e., months, not years). Exhibit 8 summarizes common methods in formative evaluation.

Exhibit 8. Common Methods in Formative Evaluation

Evaluation method	Common data sources	Notes
Implementation fidelity analysis	• Document review (e.g., service/activity logs, attendance records) • Observation forms • Case review forms • Administrative data	A study of implementation fidelity provides data to answer the first three overarching formative evaluation questions: • Is the program reaching the intended number of participants? • How are inputs contributing to program functioning? • Is the program being delivered as intended?

Evaluation method	Common data sources	Notes
Single group pre-post	• Pre-post child and family assessments • Pre-post surveys of participant experiences • Observation forms • Administrative data	A pre-post design provides client-level change data to answer the fourth overarching formative evaluation question: • Are short-term outcomes promising?
Comparison group	• Pre-post child and family assessments • Pre-post surveys of participant experiences • Administrative data	A comparison group design is often not feasible within the limited time frame for formative evaluation. However, it may be an option if you are able to easily identify and assign cases to a treatment or comparison group, or if a natural comparison group exists (e.g., an extended waitlist design).

4. Collect Data

Collect data on your selected indicators of program functioning for as long as it takes to achieve your intended sample size (as discussed in step 3).

Provide training to ensure quality data. For information on training staff in data collection, see *Critical Issues in Evaluating Child Welfare Programs* (James Bell Associates, September 2009) and *Guide to Data-Driven Decision Making* (James Bell Associates, 2018).

Approval to Collect Data

Prior to collecting data, consult with your project officer or local evaluator to determine whether you need certain approvals.

Office of Management and Budget approval may be required for federally sponsored data collection. Additional information is available from the U.S. Department of Health and Human Services (2018).

Institutional review board approval is typically required for data collection on or from individuals to ensure ethical issues are considered and privacy is maintained. Additional information is available from the U.S. Department of Health and Human Services Office for Human Research Protections (2016).

What are your response rates?

Track response rates for your data collection instruments (e.g., surveys, assessments) and identify and address issues promptly. For example, if survey response rates are low, is it because participants are given inadequate time to complete the instrument? See the Response Rate Tracking Tool (III.2).

5. Analyze and Report Data

The final phase of formative evaluation is to analyze your data, interpret the results, and clearly summarize your findings in a way key internal stakeholders can understand. Use the key formative evaluation questions in step 1 to organize your analyses and reporting. Compare the data on program indicators to the benchmarks set in step 2. Summarize whether each indicator **exceeded, met, or fell short** of the benchmark. The statistical significance of patterns in the data is not particularly important in formative evaluation—your sample size will be relatively small by design and therefore underpowered (see exhibit 1).

In reports and presentations, describe what is working and not working regarding implementation and short-term outcomes. Document lessons learned and modifications made to the program while conducting the formative evaluation.

You may adapt the Washington County Presentation of Formative Evaluation Findings to present evaluation results to stakeholders, get their feedback on a draft report, and guide decision making about next steps for your program. Also see *Developing an Effective Evaluation Report* (Centers for Disease Control and Prevention, 2013).

6. Decide What Comes Next

After formative evaluation, use what you have learned about your program to determine whether it is ready for summative evaluation.[4] It may not be. Remember, moving too quickly to summative evaluation may account in part for the low rate of success among evaluated programs (Epstein & Klerman, 2013).

[4] External constraints such as the time limitations of your funding source may also influence your decisions about whether and when to conduct summative evaluation. See the Washington County Formative Evaluation Time Line.

Summative evaluation is also costly—both financially and in terms of staff and participant time and effort. Formative evaluation allows you to pause thoughtfully before investing scarce resources in summative evaluation (see Akin et al., 2014).

The process of deciding whether and when to proceed to summative evaluation is sometimes called an evaluation tollgate. The "toll" that must be paid at the tollgate is evidence—

- That the program's logic model is plausible
- That core program components are functioning as intended
- That participants are interested and responsive
- That the program is being delivered with increasing fidelity to the model
- That short-term outcomes are trending in the right direction

If evidence is lacking in some or most of these categories, it may be premature to move to summative evaluation. Instead, use the formative evaluation findings to guide continued exploration into what is not working as planned and why, and to make improvements where needed.

The process of deciding whether and when to proceed to summative evaluation is sometimes called an evaluation tollgate. The "toll" that must be paid at the tollgate is evidence.

The problem may lie with the validity of the program—a theory problem—or with the integrity with which it is being delivered—an implementation problem (Klein & Sorra, 1996; Permanency Innovations Initiative Training and Technical Assistance Project, 2013; Testa & Poertner, 2010). If the formative evaluation shows outcomes are not moving in the expected direction, identify where the breakdown is occurring (Funnell & Rogers, 2011). Consider the following questions:

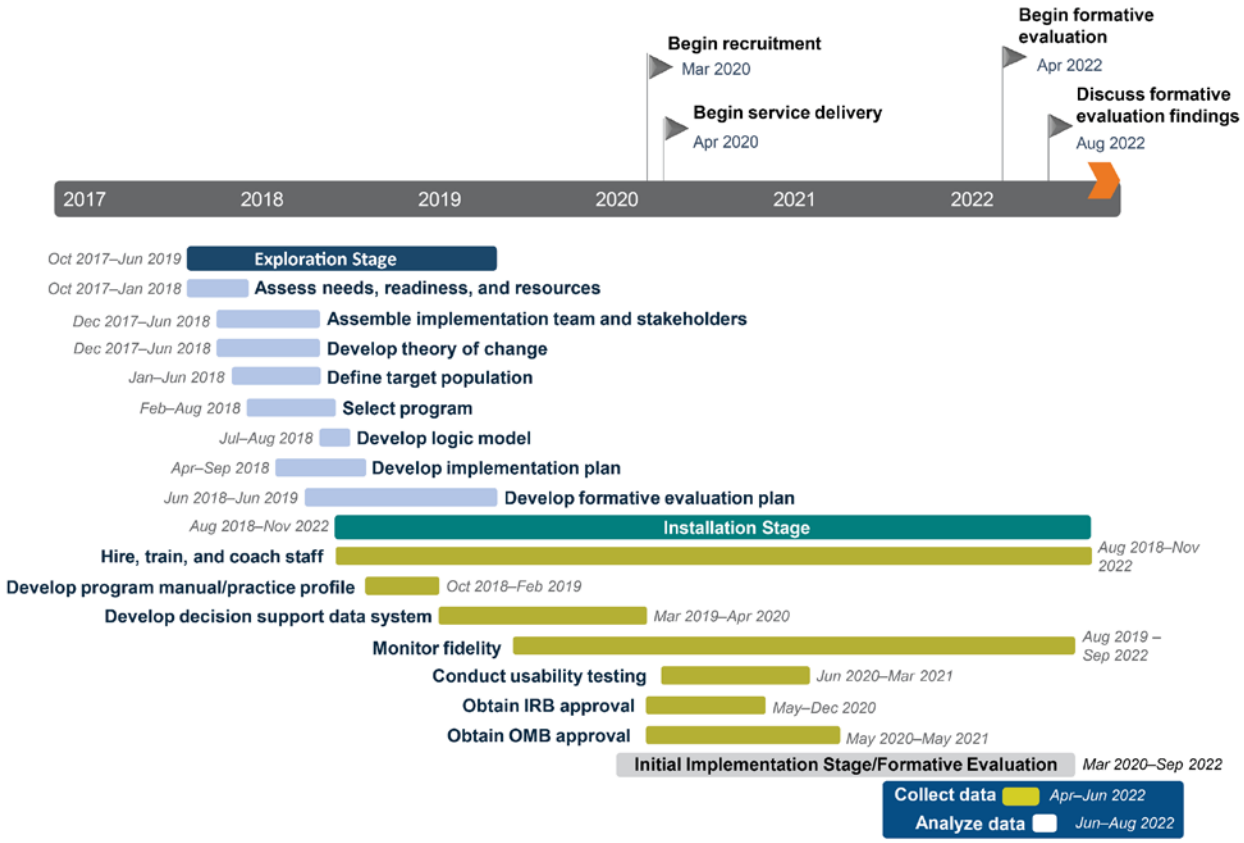
- Does your theory of change include erroneous assumptions that weaken program validity?
- Does the program lack a theory-driven approach altogether (Segal, Opie, & Dalziel, 2012)?
- Do causal mechanisms that worked with one population or context not hold true with others?
- Is the program theory sound, but elements of the program were *delivered* incorrectly, thus causing a breakdown in implementation integrity?

If you find you have either a theory problem or an implementation problem, re-examine your assumptions and/or resume the testing and improvement process (see section II).

Section III Examples

Washington County Formative Evaluation Time Line

It may take several years to reach full implementation of a new child welfare program. When planning your formative evaluation, consider how long it may take to meet the preconditions described in section II. See the example below for the fictional Washington County program.



Washington County Program Enrollment Tracking

Washington County used the table below to track enrollment in its program. See appendix A for a blank version you can use to track enrollment in your program.

Week	Number newly enrolled	Number of drop outs	Total number	Targeted total number	Difference between targeted number and total number	Number of enrolled participants	
						In target population	Outside target population
1	4	0	4	10	-6	4	0
2	6	0	10	20	-10	9	1
3	1	0	11	30	-19	11	0
4	3	0	14	40	-26	14	0
5	2	0	16	40	-24	16	0
6	0	0	16	40	-24	16	0
7	1	1	16	40	-24	16	0
8	4	0	20	40	-20	20	0
9	0	0	20	40	-20	20	0
10	5	2	23	40	-17	23	0
11	1	1	23	40	-17	23	0
12	0	0	23	40	-17	23	0
13	4	0	27	40	-13	27	0
14	2	0	29	40	-11	29	0
15	1	0	30	40	-10	30	0
16	0	0	30	40	-10	30	0

Washington County Response Rate Tracking

Washington County used the table below to track response rates for its primary data collection instruments. See appendix A for a blank version you can use to track response rates for your program.

Week	Instrument name	Number administered	Number completed	Response rate
1	Family Risk Assessment	4	4	100%
	Participant Survey	4	4	100%
	Program Notes	4	3	75%
	Case Review Forms	4	4	100%
	In-Home Visit Observation Form	1	1	100%
2	Family Risk Assessment	6	6	100%
	Participant Survey	6	4	67%
	Program Notes	6	2	34%
	Case Review Forms	10	1	10%
	In-Home Visit Observation Form	1	0	0%
3	Family Risk Assessment	1	1	100%
	Participant Survey	1	1	100%
	Program Notes	11	11	100%
	Case Review Forms	5	5	100%
	In-Home Visit Observation Form	1	1	100%
4	Family Risk Assessment	3	3	100%
	Participant Survey	3	3	100%
	Program Notes	14	13	93%
	Case Review Forms	7	5	71%
	In-Home Visit Observation Form	1	1	100%
5	Family Risk Assessment	2	2	100%
	Participant Survey	2	2	100%
	Program Notes	16	12	75%
	Case Review Forms	8	8	100%
	In-Home Visit Observation Form	1	1	100%

Week	Instrument name	Number administered	Number completed	Response rate
6	Family Risk Assessment	0	0	100%
	Participant Survey	0	0	100%
	Program Notes	16	14	88%
	Case Review Forms	8	7	88%
	In-Home Visit Observation Form	1	1	100%

Washington County Short-Term Outcome Tracking

Washington County used the table below to track short-term outcomes for its program. See appendix A for a blank version you can use to track short-term outcomes for your program.

Short-term outcome	Is it intended to increase or decrease?	Formative evaluation questions
Decrease in risk assessment ratings at program completion	Decrease	To what degree have participants' risk assessment ratings changed between program enrollment and completion?
Parents demonstrate improved parenting skills, including active listening and problem solving	Increase	What percentage of parents demonstrate improved parenting skills at program completion?
Parents demonstrate conflict management strategies at program completion	Increase	What percentage of parents demonstrate improved conflict management strategies at program completion?
Parents develop family safety plan in case of crisis situations during program enrollment	Increase	What percentage of parents have developed a family safety plan by program completion?
Fewer children entering out-of-home care during program enrollment	Decrease	What percentage of children whose parents are enrolled in the program enter out-of-home care during program enrollment?
Parents' awareness of community-based services increases during program enrollment	Increase	To what degree have parents' awareness of community-based services increased?
Child welfare staff routinely review program data	Increase	How frequently do child welfare staff review program data?

Washington County Formative Evaluation Questions and Indicators/Benchmarks

Washington County used the table below to document its research questions and indicators/benchmarks. See appendix A for a blank version you can use for your program.

Outputs	Formative evaluation questions	Indicators/benchmarks
Number of intake family assessments and number of ongoing risk assessments completed	• How many intake family assessments are completed each week?	• 85% of families with a screened-in maltreatment case complete a family assessment within 24 hours of case opening
	• How many ongoing risk assessments are completed with participating parents?	• 90% of enrolled families complete a risk assessment at program completion • 85% of enrolled families complete a risk assessment at 6 months post-program • 75% of enrolled families complete a risk assessment at 12 months post-program
Percentage of parents referred to program	• Of parents who complete the intake family assessment, what percentage are referred to the program?	• 60% of families are referred in the program
Percentage of parents engaged and enrolled within 24 hours of referral	• What percentage of parents are engaged by program staff and enrolled within 24 hours of referral?	• 100% of enrolled families are contacted by program staff within 24 hours of referral
Number and percentage of parents participating in home visits	• How many in-home visits are scheduled per week per family?	• At least two in-home visits are scheduled per week per family
	• What percentage of families completed scheduled in-home visits?	• 75% of scheduled in-home visits are successfully held
Type and dosage of services delivered to parents	• What type of services (e.g., therapeutic strategy, parenting skills, coaching) are delivered to parents by certified family therapists during in-home visits?	• At least one key service delivered as part of each in-home visit
	• How many hours of each service type are delivered per family?	• At least 4 hours of each key service delivered per family

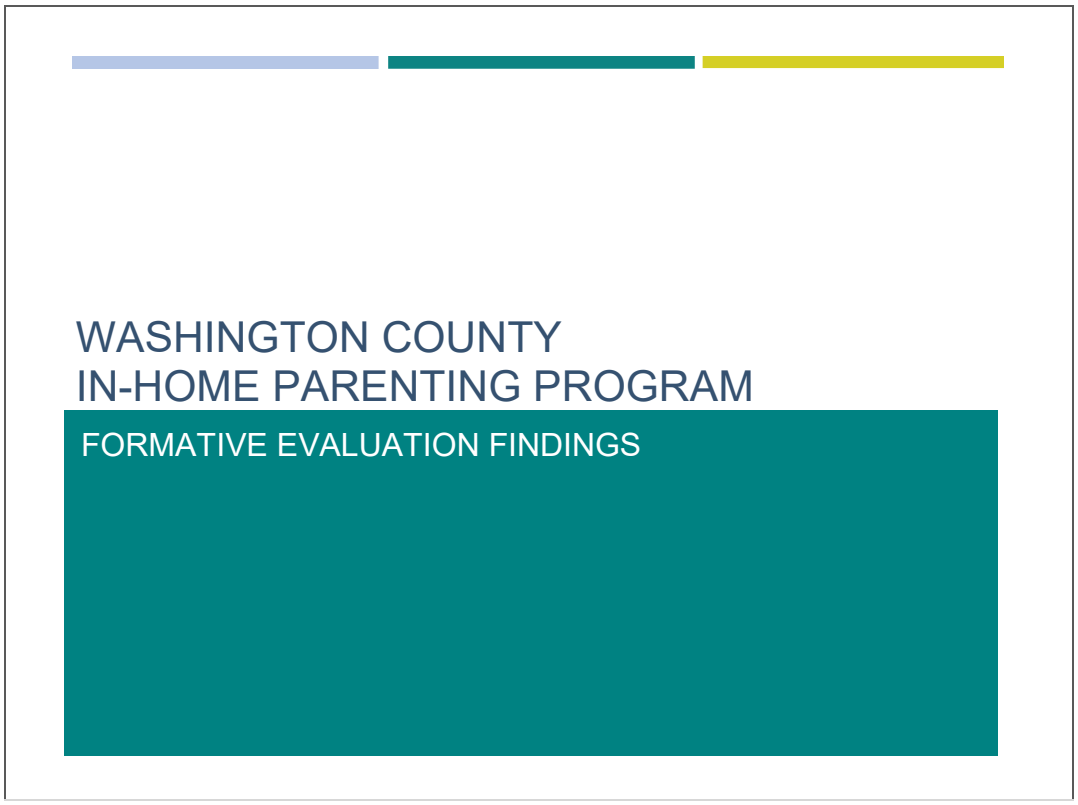
Outputs	Formative evaluation questions	Indicators/benchmarks
Frequency and type of supervisor oversight activities	• How often do supervisors review parent progress?	• Parent progress reviewed weekly by supervisors
	• What strategies (e.g., observation, secondary case note review) do supervisors use to oversee program activities?	• Supervisors observe one in-home visit per family • Supervisors review case notes weekly
Number of weekly staff meetings and attendee participation	• How many weekly staff meetings are held as scheduled?	• 90% of weekly staff meetings held as scheduled
	• What percentage of weekly staff meetings are attended by all required parties?	• 75% of weekly staff meetings attended by all required parties
Number of parent support groups held and percentage of participating families	• How many parent support groups are held during the program period?	• Parent support groups held every other week
	• What percentage of families attend parent support groups?	• 75% of families attend at least one parent support group
Number and type of therapist contacts outside of scheduled in-home visits	• How many contacts do therapists have with parents outside of scheduled in-home visits?	• At least one phone call, text, or email with parents each week
Number of accurate and complete weekly program notes	• How many program entries are completed accurately and on time?	• 90% of weekly program entries completed accurately by Monday of the following week

Short-term outcomes	Formative evaluation questions	Indicators/benchmarks
Decrease in risk assessment ratings at program completion	• To what degree have participants' risk assessment ratings changed at program completion?	• Risk assessment ratings of all enrolled families decrease by 10% or more at program completion
Parents demonstrate improved parenting skills, including active listening and problem solving	• What percentage of parents demonstrate improved parenting skills at program completion?	• 90% of parents demonstrate improved parenting skills at program completion

Short-term outcomes	Formative evaluation questions	Indicators/benchmarks
Parents demonstrate conflict management strategies at program completion	• What percentage of parents demonstrate conflict management strategies at program completion?	• 90% of parents demonstrate conflict management strategies at program completion
Parents develop family safety plan in case of crisis situations during program enrollment	• What percentage of parents have developed a family safety plan by program completion?	• 90% of parents have developed a family safety plan
Parents' awareness of community-based services increases during program enrollment	• To what degree have parents' awareness of community-based services increased?	• 90% of parents can identify recommended community-based services
Child welfare staff routinely review program data	• How frequently do child welfare staff review program data?	• Program data reviewed weekly by child welfare staff

Washington County Presentation of Formative Evaluation Findings

Use this example when developing a presentation to share your own findings with staff and key stakeholders.

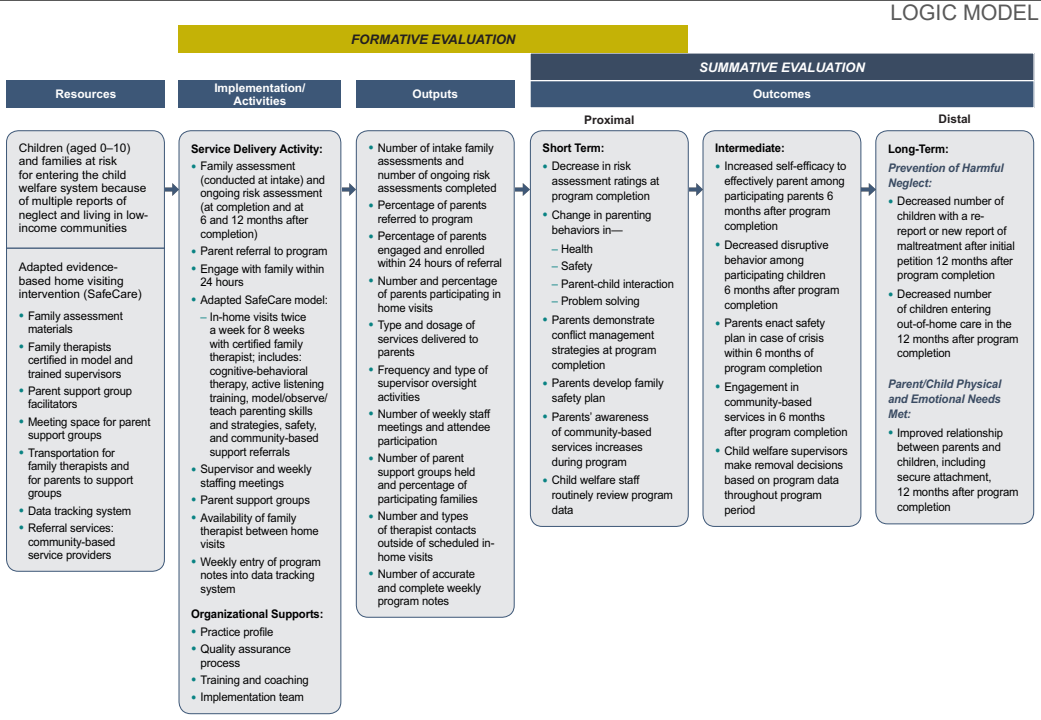


IN-HOME PARENTING PROGRAM OVERVIEW

- **Target Population:** Parents with young children (aged 0–10) at risk of removal

- **Intervention:** Intensive 8-week in-home parenting program including at least two in-home visits a week from a family therapist certified in the adapted evidence-based program. Each in-home visit lasts 1–2 hours and includes a variety of service strategies, including cognitive-behavioral therapy, parenting skills (e.g., active listening, therapists' modeling positive parenting strategies, and coaching parents in new skills), developing an in-home safety plan, and referral to additional community-based services. Parents are also encouraged to attend weekly parent support groups that are hosted by the program.

LOGIC MODEL



FORMATIVE EVALUATION QUESTIONS

- How are the program's core components functioning?
- Is the program reaching the intended number of participants?
- Is the program being delivered with fidelity?
- Are short-term outcomes promising?

FORMATIVE EVALUATION DESIGN

- Implementation Fidelity Analysis
  - Document review (e.g., service/activity logs, attendance records)
  - Observation forms
  - Case review forms
- Single Group Pre-Post Design
  - Pre-post risk assessments
  - Pre-post surveys of participant experiences
  - Observation forms

FINDINGS: CORE COMPONENT FUNCTIONING

How are the program's core components functioning?

Core component/output		Actual	Target	Finding
Risk assessment	Families with a screened-in maltreatment case completed a family assessment within 24 hours of case opening	78%	85%	Did not meet target
	Families completed risk assessment at program completion	93%	90%	Met target
	Families completed risk assessment at 6 months post-program	62%	85%	Did not meet target
	Families completed risk assessment at 12 months post-program	35%	75%	Did not meet target
	Risk assessment ratings of all enrolled families decreased by 10% or more at program completion	85%	100%	Did not meet target
Reach and exposure	See next two slides			
Child welfare operations	Accurate and complete weekly program notes	55%	90%	Did not meet target

FINDINGS: REACH

Is the in-home parenting program reaching the intended number of participants?

Reach		Actual	Target	Finding
Enrollment	Percentage of parents who completed the intake family assessment and were enrolled in the program	83%	60%	Met target
Engagement	Percentage of enrolled families contacted by program staff within 24 hours of referral	72%	100%	Did not meet target

FINDINGS: EXPOSURE

How much program content is being received by participants?

Exposure		Actual	Target	Finding
Treatment days and completion	Total days of program treatment	10	16	Did not meet target
	Number of in-home visits scheduled per week per family	2	2	Met target
	Percentage of scheduled in-home visits successfully held	76%	75%	Met target
	Number of families completing program	30	34	Did not meet target
	Number of families dropped out of program	4	0	Did not meet target
Type and dosage of services	Number of key services delivered as part of each in-home visit	1	1	Met target
	Number of hours of each key service delivered per family	4	4	Met target
	Parent support groups held every other week	Bi-weekly	Bi-weekly	Met target
	Families attend at least one parent support group	50%	75%	Did not meet target
	Number of therapist contacts outside of scheduled in-home visits	3 per week	1 per week	Met target

FINDINGS: PROGRAM FIDELITY

Is the program being delivered with fidelity?
Some aspects of program fidelity are covered in the formative evaluation questions related to core components, reach, and exposure described in previous slides.

Fidelity dimension	Finding
Adherence	See previous slides regarding core components, reach, and exposure.
Exposure	See previous slide regarding exposure.
Quality of delivery	Observation of a sample of in-home visits found that four out of five trained therapists met performance criteria.
Participant responsiveness	Participant feedback about the program on post-participation surveys was positive. Participants indicated they felt respected by the therapists and most (80%) participants reported an increase in self-efficacy to parent effectively at program completion.
Program differentiation	While no other intensive in-home parenting program is available in Washington County, parenting classes with similar material are available. Only two parents participated in the parenting classes offered outside of the program's core components. The extent to which other parents are exposed to other parenting interventions should be monitored.

FINDINGS: PROXIMAL OUTCOMES

Are short-term outcomes promising?

Short-term outcome	Actual	Target	Finding
Overall family risk assessment score	No difference in risk score	Decrease of 10%	Did not meet target
Improved parenting skills	Demonstrated by all parents	90% of parents	Met target
Conflict management strategies	Demonstrated by all parents	90% of parents	Met target
Family safety plan	Created by all parents	90% of parents	Met target
Number of children entering out-of-home care during program enrollment	75% of children did not enter out-of-home care during program enrollment	90% of children whose parents are enrolled in the program do not enter out-of-home care during program enrollment	Did not meet target
Awareness of community-based services	Demonstrated by all parents	90% of parents	Met target
Child welfare staff routinely review program data	Program data reviewed weekly by child welfare staff	Program data reviewed weekly by child welfare staff	Met target

LESSONS LEARNED

Topic	Lessons learned
Risk assessment	Additional staff training is needed to ensure all families complete a risk assessment at intake. Also, more engagement with families (calls and/or emails) after program completion may be necessary to ensure that 6- and 12-month follow-up assessments are completed.
Reach	It would be beneficial if contact by program staff could be made at the time of referral, rather than separately.
Exposure	Some families had trouble keeping their scheduled in-home visits. Total program length may need to be extended in these cases so parents can get the full program dosage. Also, parents experienced barriers to attending support groups. Online alternatives or offering groups on different days and times may be necessary.
Fidelity: program differentiation	The number of parents who participate in parenting programs or services outside the program should be carefully monitored.
Short-term outcomes	While parents demonstrated improved parenting skills at program completion, their risk scores did not change. Program theory and causal linkages should be re-examined to identify any gaps or assumptions that may be limiting program success in this area.

References

Akin, B., Testa, M. F., McDonald, T. P., Melz, H., Blase, K. A., & Barclay, A. (2014). Formative evaluation of an evidence-based intervention to reduce long-term foster care: Assessing readiness for summative evaluation. *Journal of Public Child Welfare, 8*(4), 354–374.

Centers for Disease Control and Prevention (2013). *Developing an effective evaluation report: Setting the course for effective program evaluation*. Atlanta, Georgia: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Division of Nutrition, Physical Activity and Obesity.

Deming, W. E. (1986). *Out of the crisis*. Cambridge, MA: Massachusetts Institute of Technology Center for Advanced Engineering Study.

Dhillon, L., & Vaca, S. (2018). Refining theories of change. *Journal of Multidisciplinary Evaluation, 34*(30), 64–87.

Epstein, D., & Klerman, J. A. (2013). When is a program ready for rigorous impact evaluation. The role of a falsifiable logic model. *Evaluation Review*, *36*, 375–401.

Fixsen, D., Naoom, S., Blase, K., Friedman, R., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication No. 231). Tampa, FL: Louis de la Parte Florida Mental Health Institute, University of South Florida.

Fraser, M. W., Richman, J. M., Galinsky, M. J., & Day, S. H. (2009). *Intervention research: Developing social programs*. New York, NY: Oxford University Press.

Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco, CA: Jossey-Bass/Wiley.

Geary, E., Poes, M., Iannone-Walker, M., Porter, R., Callis, A., Buckless, B., & Day, P. (2018). *Data system improvement toolkit*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Institute for Healthcare Improvement. (2018). *Plan-Do-Study-Act (PDSA) worksheet*. Retrieved from http://www.ihi.org/resources/Pages/Tools/PlanDoStudyActWorksheet.aspx

James Bell Associates. (2007, August). *Evaluation brief: Developing a logic model*. Arlington, VA: Author.

James Bell Associates. (2009, September). *Evaluation brief: Critical issues in evaluating child welfare programs*. Arlington, VA: Author.

James Bell Associates. (2009, October). *Evaluation brief: Measuring implementation fidelity*. Arlington, VA: Author.

James Bell Associates. (2018). *Guide to data-driven decision making: Using data to inform practice and policy decisions in child welfare organizations*. Washington, DC: Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services.

Klein, K. J., & Sorra, J. S. (1996). The challenge of innovation implementation. *Academy of Management Review, 21*(4), 1055–1080.

Laura and John Arnold Foundation. (2017). *How to solve U.S. social problems when most rigorous program evaluations find disappointing effects*. Retrieved from http://www.straighttalkonevidence.org/2018/03/21/how-to-solve-u-s-social-problems-when-most-rigorous-program-evaluations-find-disappointing-effects-part-one-in-a-series/

Permanency Innovations Initiative Training and Technical Assistance Project. (2013). *The PII approach: Building implementation and evaluation capacity in child welfare* (Rev. ed.). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Children's Bureau; and Office of Planning, Research and Evaluation.

Permanency Innovations Initiative Training and Technical Assistance Project. (2016). *Guide to developing, implementing, and assessing an innovation.* Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Children's Bureau.

Scriven, M. (1997). Truth and objectivity in evaluation. In E. Chelimsky & W.R. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 477–500). Thousand Oaks, CA: Sage.

Segal, L., Opie, R.S. & Dalziel, K. (2012). Theory! The Missing Link in Understanding the Performance of Neonate/Infant Home-Visiting Programs to Prevent Child Maltreatment: A Systematic Review. The Milbank Quarterly. 90(1) p. 47-106.

State Implementation and Scaling-up of Evidence-based Practices Center and National Implementation Research Network. (2017). *Topic 2: Usability testing*. Retrieved from http://implementation.fpg.unc.edu/module-5/topic-2-usability-testing.

Testa, M. F., & Poertner, J. (2010). *Fostering accountability: Using evidence to guide and improve child welfare policy*. New York, NY: Oxford University Press.

Testa, M. F., & White, K. R. (2014). Insuring the Integrity and Validity of Social Work Interventions: The Case of the Subsidized Guardianship Waiver Experiments. *Journal of Evidence-Based Social Work*, *11*(1-2), 157-172.

U.S. Department of Health & Human Services. (2018). *Information collection and Paperwork Reduction Act (PRA) overview*. Retrieved from https://www.usability.gov/how-to-and-tools/guidance/pra-overview.html

U.S. Department of Health and Human Services, Office for Human Research Protections. (2016). *Regulations & policy: Guidance.* Retrieved from https://www.hhs.gov/ohrp/regulations-and-policy/guidance/index.html

Appendix A. Tools

Tool II.1. Readiness for Formative Evaluation Checklist

Use this checklist to assess preparation for formative evaluation and to identify missing or inadequate components. It draws on the evidence-building framework and tools developed by the Children's Bureau Permanency Innovations Initiative Training and Technical Assistance Project and evaluation team.

Teaming Structure

Steps and indicators of readiness		✓	Resources/tools
Assembled implementation team(s)	Engaged agency leaders and decision makers, program staff, and supervisors who will be responsible for implementation; individuals with CQI and program data knowledge and skills; evaluators; and external stakeholders		• *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 1, Teaming Structure Assessment Tool (pages 10–12)
Documented decisions and plans	Began documenting all key decisions and developed a workplan with realistic time lines as part of an implementation and evaluation plan		• *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 2, Section 5

Problem and Target Population Identification

Steps and indicators of readiness		✓	Resources/tools
Identified the problem or outcome the team is seeking to change	Researched what the problem is (unmet need, improvement, etc.)		• *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 2, Identifying the Problem and Understanding the Target Population Tool (pages 15–17) Root Cause/Situational Analysis: • *Guide to Data-Driven Decision Making* • *Root Cause Analysis: Tracing a Problem to Its Origins*
	Identified the outcome(s) to address		
	Articulated outcomes that are clear and realistic (changeable)		

Steps and indicators of readiness		✓	Resources/tools
Identified the target population most affected	Mined or gathered data to pinpoint the affected population		• *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 2, Identifying the Problem and Understanding the Target Population Tool (pages 15–17) Root Cause/Situational Analysis: • *Guide to Data-Driven Decision Making* • *Root Cause Analysis: Tracing a Problem to Its Origins*
	Examined data longitudinally and examined geographic or disproportionate representation of the target population		
Explored needs or root causes that affect outcomes in the population	Conducted a root cause analysis to understand underlying needs		
	Mined or gathered data from case, survey, and administrative data		
	Identified systemic barriers that may affect the success of any program		

Theory of Change

Steps and indicators of readiness		✓	Resources/tools
Developed hypothesis about what activities, if implemented, will change outcomes for the target population	Developed an outcomes chain to specify hypotheses about casual links		• *Theory of Change: A Practical Tool for Action, Results, and Learning* • *Evaluation Resource Guide for Children's Bureau Discretionary Grantees* • *Guide to Data-Driven Decision Making* • *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 2, Theory of Change Development Tool (pages 27–29) • *Refining Theories of Change*
	Critically evaluated the chain and differentiated outcomes		
	Identified and addressed gaps in theory through research and data mining		
Articulated the process by which outcomes for the target population will change	Developed a brief narrative description of the theory of change		

Program Assessment and Selection

Steps and indicators of readiness		✓	Resources/tools
Researched programs and assessed the evidence base	Identified possible programs used with the population and aligned with the risk factors and needs and theory of change		• *The Hexagon: An Exploration Tool* • *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 2, Innovation Assessment and Selection Tool (pages 43–47) Program Adaptation: *Making Adaptations Tip Sheet*
Researched and documented the program—	Improves the desired outcomes		
	Was researched with the same target population		
	Was successfully replicated		
	Is ready for implementation with supporting materials from the developer		
Assessed fit of program with system or agency	Explored fit of program with the agency and system, including agency priorities, initiatives, organizational structures, and community context		
Selected program to be implemented	Reviewed and compared programs and determined best fit		
	Determined whether the program may need to be adapted		

Logic Model

Steps and indicators of readiness		✓	Resources/tools
Developed a logic model and evaluated its plausibility	Drafted a logic model with resources, inputs, assumptions, activities, outputs, proximal and distal outcomes, and alignment with the organization's values		• *Developing a Logic Model* • *Guide to Data-Driven Decision Making*
	Critically evaluated the logic model, looking for implausible elements or gaps in logic		
	Set short-term benchmarks that could be tested during usability testing and formative evaluation to inform decisions regarding summative evaluation		

Program Operationalization

Steps and indicators of readiness		✓	Resources/tools
Operationalized the program based on the context of the organization and target population's needs	Team has identified the essential functions (i.e., activities or strategies that a practitioner engages in to address the identified problem) that are needed to deliver the innovation according to the model		• *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 3, Practice Profile Development Tool (pages 12–15) • *Making Adaptations Tip Sheet*
	Team has developed a practice profile or program manual that outlines— – Essential functions – The operationalized definition of each function – Activities staff need to perform to meet each function – Behaviorally based practice indicators – Practice criteria that describe different levels of performance		
	If adaptation is required, team has fully adapted the program or is working with the developer to adapt the program to fit needs		

Implementation Supports

Steps and indicators of readiness		✓	Resources/tools
Assessed and strengthened implementation supports	Assessed implementation supports and determined needs		• *Implementation Drivers: Assessing Best Practices* • *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 2, Section 4 • *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 3, Sections 7.2–7.7 Fidelity Assessment: • *Measuring Implementation Fidelity* • *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 3, Fidelity Assessment Tool (pages 40–44) and Fidelity Assessment Protocol Tool (pages 50–54)
	Ensured strong leaders communicate goals and support the initiative		
	Installed processes related to structural needs such as referral pathways, securing resources, and purchasing new technology or equipment		
Planned and implemented hiring, training, supervision, and coaching	Ensured competent staff are recruited and selected to implement the program		
	Developed protocols for training, supervision, and coaching of staff that impart knowledge and core competencies in the program		
	Developed individual and group coaching strategies and a service delivery plan		
Planned and implemented the fidelity assessment	Identified a set of key indicators to assess whether practitioners are implementing the program as intended		
	Used the practice profile to develop the fidelity assessment		
	Planned to gather and analyze fidelity data		

Data Sources and Systems

Steps and indicators of readiness		✓	Resources/tools
Data systems	Specified the organization's information needs as the program is put into place and throughout implementation		• *Guide to Data-Driven Decision Making* • *Decision Support Data System* • *Decision Support Data System Best Practices*
	Identified existing data sources and planned and installed new data sources as needed		
	Actively involved the data staff and experts to confirm quality and consistency of the data and improve the usefulness of subsequent reports		
	Developed a continuous quality improvement processes for decision making		

Usability Testing

Steps and indicators of readiness		✓	Resources/tools
Launched the program	Participants are being enrolled in the program or the innovation is launched		• *Rapid-Cycle Problem Solving* • *Usability Testing* • *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 4, Usability Testing Tool (pages 10–27) • *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 4, Appendix

Steps and indicators of readiness		✓	Resources/tools
Planned and conducted rapid-cycle testing	Team is familiar with the rapid-cycle problem solving approach (Plan-Do-Study-Act framework)		• *Rapid-Cycle Problem Solving* • *Usability Testing* • *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 4, Usability Testing Tool (pages 10–27) • *Guide to Developing, Implementing, and Assessing an Innovation*: Volume 4, Appendix
	Determined components to be tested and questions to be answered in usability tests		
	Determined methods of data collection and analysis		
	Developed a plan that outlines roles, scope of testing, identification of cases or providers, metrics and benchmarks to meet, time line for reporting, and criteria for ending or repeating the test cycle		
	Completed a full cycle(s) of testing		
	Developed a report that interprets findings and provides recommendations for actionable improvements		
	Refined the program or innovation based on findings of testing or multiple cycles of testing		
	Reviewed results and confirmed the program and implementation supports are ready for formative evaluation		

Tool II.2. Refining Your Theory of Change Tool

Use this tool to identify areas for refinement in your theory of change. See the Washington County Theory of Change example at the end of section II.

Topic	Review question	✓
Data exploration	Have you conducted a root cause analysis (or situational analysis)?	
	Does your theory of change reflect the findings from the root cause analysis, such as extent of the problem, population most affected, risk characteristics, and so on?	
	Does your theory of change reflect the conditions that need to change to achieve the desired outcome?	
Theory of change components	Does your theory of change include an outcomes chain that articulates logical and temporal relationships between actions and outcomes?	
	Does the theory of change differentiate between proximal outcomes (shorter term, which can be reasonably affected by the project activities); intermediate outcomes (can't happen without the proximal outcomes); and distal outcomes (longer term, which may not be measurable by the project or within the project's time frame)?	
	Does the theory of change clearly explain the mechanisms of change (i.e., the "why" behind how actions will trigger the expected outcomes)?	
	Examine the chain in your theory of change both forward and backward. Do the logic and casual links "hold up" in both directions? Are there any gaps in logic or tenuous assumptions?	
	Are there any gaps in the logic or flow of the theory of change—for example, activities that are "outcome-less" (i.e., not associated with one or more outcomes)?	
	Has the team considered whether there are any unintended or indirect effects that may occur and should be represented?	
	Has your theory of change been collaboratively developed or have stakeholders had opportunities to provide feedback?	
Narrative theory of change	Has the outcomes chain or pathways of change been summarized in a narrative paragraph that is easily understood by internal stakeholders and external audiences?	

Tool II.3. Refining Your Logic Model Tool

Use this tool to evaluate the plausibility and quality of your logic model and consider methods to address areas that may require more attention. See the Washington County Logic Model example at the end of section II.

Topic	Review question	✓	Methods to address
Components of your logic model	Does your model specify inputs, activities, and outputs?		Ensure all key resources and actions (processes) of the program are included. Outputs should be specified with a measurable benchmark.
	Does your model include external conditions and assumptions?		Consider and indicate how and why the program matches the needs, target population, and risk context; how it is expected to achieve outcomes; and what may affect the problem and program as it is implemented.
	Does the model specify proximal outcomes (shorter term, which can be reasonably affected by the project activities); intermediate outcomes (can't happen without the proximal outcomes); and distal outcomes (longer term, which may not be measurable by the project or within the project time frame)?		If not already completed as part of your theory of change development process, develop an *outcomes chain* by listing all potential outcomes, sorting them temporally, and examining logical linkages between outcomes and the potential for change as a result of the program.
	Does each output and outcome have a clear and well-specified indicator and benchmark?		Review your indicators to ensure they are SMART: specific, measurable, attainable, relevant, and time bound.
	Does the model reflect the "end-values" of the program, including the mission and vision of leadership and the organization?		Specify the values that will help all involved staff and systems stay engaged and motivated over the course of implementation ("why we do this work").

Topic	Review question	✓	Methods to address
Model linkages	Examine each chain in your logic model both forward and backward. Does the logic "hold up" in both directions? Are there any gaps in logic or tenuous assumptions?		You can apply the "if-then" strategy to examine all linkages (e.g., "if X activity is implemented, then will Y short-term outcome occur?").
	Does the temporal order of activities and outcomes make sense? What are the approximate time frames for implementing each activity and achieving each outcome?		Create a table that indicates the order and time frame for implementing each activity and associated outcome.
	Are there activities in the logic model that are unnecessary or beyond the scope of the program?		Confirm the activities are part of the program's core components; remove any activities that are not core to the program and consistent with your theory of change.
Testing common areas of failure of the logic model	Could you experience input issues, such as staff turnover?		Ensure appropriate partnerships and detail staff qualifications and methods of retention.
	Could enrollment in the program be lower than expected?		Review data and monitor patterns, set realistic expectations, and install methods to address recruitment challenges.
	Are all activities logically linked to outputs and outcomes?		Ensure there are no "outcome-less" activities or large gaps in logic regarding how an activity will lead to a given outcome.
	Is the anticipated program completion rate feasible?		Specify what defines "enrollment." What proportion of attendees needs to participate in what number of activities or sessions to detect measurable effects?
	Have clear measures of fidelity to the program been developed and are they possible to achieve?		State what constitutes sufficient implementation fidelity in specific and measurable terms.
	Will you see improvement in your outcomes post-program?		Ensure proximal outcomes can be changed by the program and carefully choose methods of measuring them. Confirm they are feasible to achieve within specific time frames.

Note: Testing content adapted from Epstein & Klerman, 2013.

Tool II.4. Usability Testing Tool

Use this tool to plan for usability testing. See the Washington County Usability Testing Plan and Results at the end of section II.

Usability test(s)	Area	Usability question	Scope of testing	Participants	Roles	Metrics or key outputs	Data sources and methods	Time line	Criteria for completion
	Which area are you targeting?	*What are you trying to learn?*	*Which process or step is being tested?*	*Who will be included in the test and how will they be identified or included?*	*Who are the—* • *Decision makers?* • *Test coordinators?* • *Data collectors?* • *Analysts and reporters?*	*What are the outputs that could be assessed in the short term to answer the questions?*	*What data will be collected to measure the metric? What methods will be utilized?*	*When will the test begin and end?*	*What are the criteria for success? What is the benchmark?*
#1									
#2									
#3									
#4									

Tool III.1. Program Enrollment Tracking Tool

Use this template to track program enrollment. See the Washington County Program Enrollment Tracking example at the end of section III.

Week	Number newly enrolled	Number of drop outs	Total number	Targeted total number	Difference between targeted number and total number	Number of enrolled participants	
						In target population	Outside target population
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							

Tool III.2. Response Rate Tracking Tool

Use this template to track response rates to primary data collection instruments. See the Washington County Response Rate Tracking example at the end of section III.

Week	Instrument name	Number administered	Number completed	Response rate
1				
2				
3				

Week	Instrument name	Number administered	Number completed	Response rate
4				
5				

Tool III.3. Short-Term Outcome Tracking Template

Use this template to track short-term outcomes. See the Washington County Short-Term Outcome Tracking example at the end of section III.

Short-term outcome	Is it intended to increase or decrease?	Formative evaluation questions *Has it increased or decreased so far (i.e., is it better or worse)?*

Tool III.4. Formative Evaluation Questions and Indicators/Benchmarks

Use the table below to document research questions and indicators/benchmarks. See the Washington County Formative Evaluation Questions and Indicators/Benchmarks example at the end of section III.

Outputs	Formative evaluation questions	Indicators/benchmarks
Short-term outcomes	**Formative evaluation questions**	**Indicators/benchmarks**

Appendix B. Glossary

Adaptation	The process of modifying a program to meet local characteristics (e.g., risk and contextual factors, availability of the target population) without changing its core components. For example, an evidence-based trauma therapy program from the juvenile justice system may be adapted to work in a congregate care child welfare setting.
Causal linkages	A series of steps considered necessary for expected program outcomes to be achieved (Permanency Innovations Initiative Training and Technical Assistance Project, 2016); an explanation of the mechanisms through which programs or activities will change outcomes.
Core components	Activities or features of an intervention that define and distinguish it from other services or activities and must be present for the intervention to occur.
Decision support data system	A system (usually computerized) for collecting and organizing information about implementation processes and program outputs and outcomes. The system should provide data in a timely, reliable, and accessible manner to support decision making (Permanency Innovations Initiative Training and Technical Assistance Project, 2016).
Evaluability	The degree to which a program is defined and understood by staff, leadership, and stakeholders and can plausibly and feasibly achieve its objectives. Evaluability indicates the program is suitable for evaluation.
Evidence-based program	A program is classified as evidence based when it has undergone one or more rigorous evaluations demonstrating its effectiveness (i.e., observed positive outcomes in the target population).
Fidelity	The extent to which a program or process is implemented as designed or intended. For example, in the context of a training program, fidelity refers to the extent to which the people conducting the training follow the content, guidelines, and tools (e.g., scripts, learning exercises) provided in a training curriculum.
Implementation supports	Resources and processes that improve the likelihood of successful implementation of a program or innovation. Domains of supports include staff selection, training, coaching, fidelity assessments, identification of data and data systems, and leadership and stakeholder backing of the program or innovation.
Implementation team	A group of individuals who guide the installation and execution of a program by ensuring it “is defined, operationalized, and implemented; ensures implementation supports are in place; identified the measures for monitoring the intervention; and plans for sustaining the improved outcome” (Permanency Innovations Initiative Training and Technical Assistance Project, 2016, volume 1., p. 4).
Indicators	Clearly defined practitioner behaviors and activities used to guide training, coaching, and assessment of fidelity to the program (Permanency Innovations Initiative Training and Technical Assistance Project, 2016).
Logic model	A visual depiction of the program’s theory of change, illustrating how inputs and activities lead to outputs and outcomes.

Manualized	Written guidance that describes in detail how core components are to be implemented. For example, a program may include standardized manuals for staff and supervisors but may require tailoring to use terminology appropriate and familiar to staff or to fit within local operating practices, policies, and regulations.
Operationalize	To identify and define the activities, behaviors, and context of core components so they can be measured or assessed (Permanency Innovations Initiative Training and Technical Assistance Project, 2016).
Practice profile	A written description of a program's core components that facilitates consistent implementation and fidelity across service delivery staff (Permanency Innovations Initiative Training and Technical Assistance Project, 2016).
Rapid-cycle improvement processes	The use of observation and program data to measure changes in implementation facilitating the identification and correction of challenges. Activities are usually completed on a much shorter schedule than for a formal evaluation, typically occurring over a period of 1–3 months.
Root cause	The factor or source underlying observable characteristics that define the problem or symptoms of the problem experienced by the target population.
Target population	The group of individuals targeted by the program because they are currently experiencing a defined problem or are at risk of experiencing the problem (Permanency Innovations Initiative Training and Technical Assistance Project, 2016).
Teaming structures	The way in which a group of individuals is organized to facilitate communication and program implementation. For example, a program may have several teams responsible for different aspects of implementation, such as a management team, design team, training and coaching team, and participant recruitment team.
Theory of change	A written description or visual depiction of the assumptions about how and why a program will lead to desired outcomes. A theory of change may begin with describing the root cause of the problem and target population and then describe how the program will effect change within the target population, leading to desired outcomes.
Usability testing	The process of testing the core components of a program and data collection to identify challenges and allow for adjustments prior to entering a period of evaluation.