

Evidence-Building Evaluation Strategies for Untested and Promising Child Welfare Programs

Introduction

Child welfare agencies and service providers have placed increased emphasis in recent years on the selection and implementation of evidence-based programs and practices for children and families served by the child welfare system. At the federal level, the Research and Evaluation Workgroups established by the Children’s Bureau within the U.S. Department of Health and Human Services (HHS) articulated expectations for greater rigor and research quality in child welfare in a framework for designing, testing, disseminating, and sustaining effective practices (Framework Workgroup, 2014). This emphasis intensified with the passage in 2018 of the Family First Prevention Services Act (FFPSA), which provides optional and time-limited title IV-E funding to state and tribal child welfare agencies for certain foster care prevention programs and services that are designated as “promising,” “supported,” or “well-supported.” These designations are based on practice requirements and quality of research criteria set forth in the legislation.¹ The Act also specifies that states and tribes must conduct rigorous evaluations of funded prevention programs.²

The need to expand the child welfare evidence base remains great. A recent review of the [California Evidence-Based Clearinghouse for Child Welfare](#) (CEBC) found only 22 programs with both high relevance to child welfare and “well-supported” or “supported” evidence of positive impact.³ Despite growing interest in and demand for evidence of effectiveness, child welfare organizations face many hurdles in building evidence for untested and promising programs and services; however, more resources than ever before, including child welfare evidence clearinghouses, are available to help organizations and service providers collect and document evidence of the benefits of programs and practices for vulnerable children and families. This brief highlights the key elements and evaluation

¹ For a complete summary of the Family First Prevention Act, see Administration on Children, Youth, and Families Information Memorandum [ACYF-CB-IM-18-02](#).

² HHS may waive the evaluation requirement if it deems the evidence of the effectiveness of a given program or practice to be compelling and the child welfare agency meets the continuous quality improvement requirements for the program or practice.

³ Review conducted in May 2021.

criteria of clearinghouses that are most relevant to child welfare; factors to weigh in deciding whether to submit a program or practice for an evidence review; evaluation design considerations to build evidence; how to determine organizational capacity to conduct a rigorous evaluation; and initial planning steps for implementing and evaluating a program.

Evidence Clearinghouses: Definitions and Requirements

Evidence clearinghouses are online databases that review, rate, and catalogue the effectiveness of various social and educational programs and practices. Examples of recently developed and implemented clearinghouses include the following:

- [What Works Clearinghouse](#), established by the Institute of Education Sciences in the U.S. Department of Education, is the primary source of information on evidence-based programs and practices in K-12 education.
- [Model Programs Guide](#), sponsored by the Office of Juvenile Justice and Delinquency Prevention within the U.S. Department of Justice, presents information on the effectiveness of juvenile justice and youth delinquency prevention programs and practices.
- [National Registry of Evidence-Based Programs and Practices](#), sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA) within HHS, was the main information resource on evidence-based programs and practices in the mental health and substance abuse fields. The federal government suspended the registry in December 2017; although no new reviews of evidence-based programs are being conducted, SAMHSA continues to support the Evidence-Based Practices Resource Center through the registry's former website.

In addition to these clearinghouses, there are two main sources of information on effective programs and practices for child welfare populations and practice settings: the [Title IV-E Prevention Services Clearinghouse](#) and the [CEBC](#). These two clearinghouses are the foundation for this brief's subsequent discussions of considerations in building evidence of effectiveness in child welfare contexts.

Title IV-E Prevention Services Clearinghouse

Established by the Administration for Children and Families, HHS, under the provisions of FFPSA, the Title IV-E Prevention Services Clearinghouse seeks to review and rate research on programs focused on providing enhanced support for children and families and preventing placement of children into foster care. Four program and service areas are eligible for review and rating:

- Mental Health Prevention and Treatment Programs and Services
- Substance Abuse Prevention and Treatment Programs and Services
- In-Home Parent Skill-Based Programs and Services
- Kinship Navigator Programs

As codified in FFPSA, the clearinghouse assigns reviewed programs and services to one of the following ratings categories: “well-supported,” “supported,” “promising,” or “does not currently meet criteria.” For complete information on submission requirements, the review process, and definitions of ratings, see the [Title IV-E Prevention Services Clearinghouse Handbook of Standards and Procedures](#).⁴

California Evidence-Based Clearinghouse

The CEBC is funded by the California Department of Social Services Office of Child Abuse Prevention and is operated and maintained by the Chadwick Center for Children and Families at Rady Children’s Hospital–San Diego. The CEBC’s purpose is to build the knowledge base for and advance the implementation of evidence-based practices for children and families involved in the child welfare system. The CEBC classifies programs by topic area, child age, and delivery setting and includes two ratings scales for level of evidence and relevance to child welfare. The evidence categories include “well-supported by research evidence,” “supported by research evidence,” “promising research evidence,” “evidence fails to demonstrate effect,” “concerning practice,” and “not able to be rated”; the child welfare relevance ratings include “high,” “medium,” and “low.” For complete information on CEBC submission requirements and definitions of ratings, see the [CEBC Rating Policy and Procedures Manual](#).

Exhibit 1 describes the minimum eligibility criteria these two clearinghouses use to determine whether a given program and any published studies of the program are suitable for review and inclusion in their program databases.

⁴ HHS recently published a [Federal Register Notice](#) to request public feedback on the *Handbook of Standards and Procedures*, specifically with respect to the identification of programs for review, the selection and prioritization of programs for review, study eligibility screening and prioritization criteria, program evidence ratings, and evidence review procedures.

Exhibit 1. Eligibility Criteria for Review by Child Welfare Clearinghouses

Criteria	Title IV-E Prevention Services Clearinghouse	CEBC
Scope	Priority given to programs/services that are in use, target outcomes in eligible domains (see Outcomes category below) and have implementation supports in place (including fidelity tools, training, and coaching)	Examines the intervention in its entirety, not just one or more components of the intervention
Manualization	Must have publicly available written protocols, manuals, or other documentation describing how to implement/administer the program or practice ⁵	Intervention/program in a reviewed study must be conducted in its manualized form
Research design	Studies must use a randomized control trial (RCT) or quasi-experimental group design and include at least one contrast (defined as a comparison of a treated condition to a counterfactual condition on a specific outcome). Comparison groups must be “no or minimal intervention” or “treatment as usual”	Research design must use some form of control or comparison group. The control or comparison may include an RCT control group, a waitlist control group, an untreated group, or other similar type of control/comparison
Outcomes	Target outcomes must be in the domains of child safety, child permanency, child well-being, and/or adult well-being; kinship navigator outcomes must be in the domains of child safety, child permanency, child well-being, adult well-being, access to services, referral to services, and/or satisfaction with programs and services	No required/targeted child, caregiver, or family outcomes specified; however, the clearinghouse’s relevance ratings are used to identify programs that are designed for or commonly used to meet the needs of children, youth, or families receiving child welfare services
Publication/ dissemination	Reviewed studies must be publicly available; in English; and published in or after 1990 in peer-reviewed journals or reports prepared or commissioned by federal, state, or local government agencies or departments; research institutes, research firms, or foundations; or other similar organizations. Dissertations, theses, and conference papers are not eligible	Reviewed studies must appear in published, peer-reviewed literature

Source: Title IV-E Prevention Services Clearinghouse and CEBC

⁵ When multiple versions of a program or service are available, the Prevention Services Clearinghouse selects just one version for review at a time and reviews eligible studies only of the version selected. Other versions may be eligible for review as separate programs or services.

Obtaining a Clearinghouse Rating: Thinking Through the “Why’s” of Implementing and Evaluating Your Program

A child welfare organization seeking to build evidence for an untested or promising intervention may wish to focus its efforts on obtaining a review and rating from the Prevention Services Clearinghouse or CEBC given their direct relevance to child welfare programs and service populations. However, considerable time and resources are required to design and implement a research study or program evaluation that is sufficiently rigorous to meet the rating criteria of these clearinghouses—and even more resources are required to obtain their highest evidence ratings. First carefully think about your organization’s goals, objectives, and reasons for seeking an evidence rating. Examples include the following:

- Obtaining more access to grant opportunities from federal, state, or local government agencies, nonprofits, or foundations
- Increasing organizational revenue more generally (e.g., gaining access to title IV-E funds under the provisions of FFPSA)
- Expanding or scaling up a program within your organization or within other agencies or service systems
- Making more programs/services available to certain target populations your organization serves
- Building organizational and/or evaluation capacity within your organization

Your organization should also carefully consider the pros, cons, and trade-offs of investing time and effort in seeking a clearinghouse evidence rating. Reflect on the following questions with key decision makers in your organization:

- What commitment of organizational time, personnel, and resources will be required to implement a research study or evaluation that meets the eligibility requirements of one or more clearinghouses? What resources might be diverted from other programs, services, operations, evaluation activities, etc., in seeking an evidence rating?
- What level of evidence rating do we wish to obtain? Are we striving for the highest rating (i.e., “well-supported” in both clearinghouses)? Or is a lower rating (e.g., “supported,” “promising”) acceptable?
- What are the risks of failure? What might happen if an evaluation of our program does not show conclusive evidence of benefit and receives a lower evidence rating from a clearinghouse, or is rejected for consideration altogether (e.g., bad publicity, loss of funding or advocacy and support from your organization)?
- Is a high evidence rating itself adequate to realize our organization’s goals in seeking the rating (e.g., access to more funding, expanding the program to new populations or service

settings)? Would the achievement of these goals justify the investment of time and resources that would be necessary to obtain a high rating, or would the pursuit of a lower rating that requires a smaller investment but still demonstrates some evidence of effectiveness be a better option?

Steps in Building Evidence

If your organization decides to invest resources in a research study or evaluation to make your program eligible for clearinghouse review, some preliminary efforts are necessary prior to submitting your program for consideration. This section lists the steps that will ensure your program is ready for clearinghouse review and assessment.

Assessing Evaluation Capacity and Initial Planning

Preparing for the evaluation of any program first requires an assessment of your organization's resources and capacity to conduct or sponsor a high-quality evaluation effort. For example, does your organization have an internal research and evaluation unit, or will you need to partner with an external evaluator? There are several important questions to consider when determining who will conduct your evaluation; for additional guidance, see JBA's brief on [Selecting and Working With an Evaluation Partner](#).

After engaging an external evaluator or confirming your organization's internal capacity to conduct an evaluation, initial preparations include gathering data to inform the selection of an evaluation approach that will most accurately measure your program's outcomes and build evidence of effectiveness. Key variables to explore during initial evaluation planning include the following:

- Population characteristics: What are the demographic characteristics and needs of the children and families, and of the broader community, in which the program will be implemented?
- Population size: How large is the group of children, caregivers, or families your program will serve? Will you seek to serve the entire population of potentially eligible persons or just a sample from that population? Too small of a sample may not provide enough statistical power to detect meaningful differences in outcomes between program participants and a control or comparison group of nonparticipants.
- Current services: What services, treatments, and supports are currently available in your target community or in communities or populations that may serve as a comparison group? How similar are they to the program you plan to evaluate? Knowing and controlling for the kinds of services that are already provided in your target and comparison communities can help minimize design contamination (i.e., the dilution of your program's impacts due to the presence and use of similar interventions).

- Stakeholder engagement and program champions: Do you have a plan to communicate with and engage potential program participants, staff within your organization, and other stakeholders in the community to increase their interest and buy-in? Have you identified and trained staff or community members who can serve as “champions” to lead messaging, recruitment, and retention efforts?

Conducting a Formative Evaluation

The next step is to assess your program’s readiness for a methodologically rigorous and comprehensive evaluation of its impacts on targeted child and family outcomes. This type of evaluation, sometimes referred to as a *summative evaluation*, is usually only implemented once a program is fully operational and no major changes to program services or activities are expected (James Bell Associates, 2018). As a preliminary but important step, many organizations first choose to implement a *formative evaluation*, which is conducted during initial program planning and implementation and there is interest in or a need for continued development or refinement of program activities and services. A formative evaluation provides an opportunity to examine whether a program is being implemented as originally intended and whether preliminary outcomes are promising and trending in the expected direction. It may also reveal unexpected or unintended findings, which gives leadership an opportunity to make additional program improvements.

Before you undertake a formative evaluation, several preparatory activities are necessary to ensure your program is the right fit for the problem(s) you seek to address and the populations you intend to serve. Key activities to complete in preparation for program implementation and formative evaluation include the following:

- Problem exploration and root cause analysis: What is the problem you are attempting to solve, and what is the root cause of the problem?
- Theory of change development: How is the program supposed to work? How will it address the problem and achieve desired outcomes?
- Program operationalization: Does your program have clear policies, practices, and protocols in place? Have they been clearly outlined in a detailed program manual and staff training curriculum?
- Identification of required data and installing data collection systems: Have you determined what information you will need to conduct a formative evaluation? Have you identified the sources of these data and/or developed the systems you’ll need to collect, analyze, and report the data?

Once you have completed these preparatory activities, the next steps in completing a formative evaluation include articulating your evaluation research questions, selecting indicators of program functioning (i.e., outputs and short-term outcomes, any corresponding performance targets or

benchmarks), selecting an appropriate research design (often a single group pre-/posttest or comparison group design), and collecting and analyzing data. For more specific guidance on conducting a formative evaluation, see the [Formative Evaluation Toolkit: A Step-by-Step Guide and Resources for Evaluating Program Implementation and Early Outcomes](#).

A formative evaluation may be implemented once, or it may involve several iterations to ensure your program is being implemented as intended and is demonstrating initial evidence of effectiveness based on positive trends in short-term outcomes. Eventually, your program team and other stakeholders will need to assess the evidence for several factors to determine whether the program is ready for a larger-scale summative evaluation. Specifically, they will need to establish whether—

- The program’s theory of change and logic model are plausible based on observed implementation findings and early outcomes.
- Core program components are functioning as intended.
- Program participants are engaged and receiving the expected dosage of treatment or service.
- The program is being delivered with fidelity to the program model.
- Short-term outcomes are trending in the right direction.

The process of deciding whether and when to proceed to summative evaluation is sometimes called an evaluation “tollgate” (Epstein & Klerman, 2013). If evidence is lacking in some or most of the categories noted above, it may be premature to move to summative evaluation, and more formative evaluation work may be necessary to identify areas in which improvements in program services, activities, or operations are warranted.

Preparing for Summative Evaluation: Selecting a Research Design

Once results from your formative evaluation suggest that your program is ready for a comprehensive summative evaluation or research study, selecting a research design or methodology is the next consequential step; among other factors, the methodology you choose has a direct bearing on the clearinghouse evidence rating for which your program will be eligible. Exhibit 2 highlights several design alternatives for a comprehensive research study or summative evaluation and indicates the highest level of evidence rating each design option potentially qualifies for in the Prevention Services and CEBC clearinghouses.

Programs that have been evaluated using one of several methodologies that are common in child welfare practice settings—including longitudinal/time series, pre-/posttest, and regression discontinuity designs that have no comparison group or condition (e.g., a single-group pre-posttest)—are currently ineligible to receive an evidence rating from these clearinghouses. This exclusion necessitates selecting from a more limited set of research methodologies, specifically one

Exhibit 2. Research Design Options

Design type	Description	When to use	What's needed to implement	Disadvantages/limitations	Highest possible clearinghouse rating
Randomized controlled trial (RCT)	Cases are randomly assigned to an experimental group (eligible for the new program or service) or a control group (eligible for existing/usual services).	<ul style="list-style-type: none"> • There is a discrete and clearly defined treatment or service • There are more people eligible for the service than there are resources available to provide the service 	<ul style="list-style-type: none"> • Buy-in from program management and staff • High degree of technical expertise • Strict control over the assignment process to prevent design contamination 	<ul style="list-style-type: none"> • Results may not be generalizable to other groups/settings outside of RCT population (limited external validity) 	<ul style="list-style-type: none"> • Prevention Services: Well-Supported • CEBC: Well-Supported
Waitlist/overflow	Cases are placed in a “waitlist” (comparison) group when caseloads for a new program are full. Outcomes for cases in the waitlisted group are compared with outcomes for cases that receive the program	<ul style="list-style-type: none"> • There is a discrete and clearly defined treatment or service • There are more people eligible for the service than there are resources available to provide the service • Treatment or service is of limited duration 	<ul style="list-style-type: none"> • Ability to strictly control the assignment process (assignment should be based strictly on a “first come, first served” rule) 	<ul style="list-style-type: none"> • Usually only feasible with short interventions; data collection must be completed before waitlisted cases receive the intervention • There is a higher risk of selection bias than with RCTs 	<ul style="list-style-type: none"> • Prevention Services: Well-Supported (follow-up data collection must occur at least 12 months post-intervention) • CEBC: Promising
Matched case	Each experimental group case is individually matched with a comparison case based on selected matching variables. Propensity score matching is one example of this approach	<ul style="list-style-type: none"> • The program/service targets a limited and well-defined population • The service must be provided to as many people as possible (e.g., because of small sample sizes, requirements of a service contract) • A group of matching cases with very similar characteristics can be identified 	<ul style="list-style-type: none"> • Comprehensive and detailed data are available on the case characteristics and demographics of both treatment and matching cases 	<ul style="list-style-type: none"> • Risk of bias because matching only controls for observed (and observable) variables; other factors may explain differences between groups • Biases such as regression toward the mean can occur if characteristics of the two groups do not overlap sufficiently 	<ul style="list-style-type: none"> • Prevention Services: Well-Supported • CEBC: Promising
Comparison group/site	A group or community (e.g., neighborhood, city, county) is identified with characteristics similar to those of the target group/community in which the program is implemented; differences in outcomes between the groups are compared over time	<ul style="list-style-type: none"> • A system- or community-wide reform effort/initiative is being implemented rather than a discrete program, treatment, or service • Complete and accurate case-level data prior to implementation are not available • A community exists that has similar characteristics <i>and</i> no similar services or reform efforts in place 	<ul style="list-style-type: none"> • Cooperation and assistance from authorities in other jurisdictions around data sharing and collection • Ability to collect baseline data from both treatment and comparison groups/communities 	<ul style="list-style-type: none"> • High risk of selection bias; different outcomes may result from unmeasured differences between the groups • Causal relationships between the intervention and outcomes cannot be firmly established due to extraneous and uncontrollable confounding variables 	<ul style="list-style-type: none"> • Prevention Services: Well-Supported (requires at least two intervention sites/communities and two comparison sites/communities with no overlap) • CEBC: Promising

with a clearly defined treatment and control/comparison condition. Whereas both clearinghouses designate randomized controlled trials (RCTs) as the most rigorous design option, the Prevention Services Clearinghouse does consider a wider range of quasi-experimental designs for its highest program and service rating—provided their associated studies meet several quality standards with respect to sample size, baseline equivalence of samples, and controls for potential confounding variables, among other factors.

Another factor to consider is where on the evidence continuum your program currently lies; if it is a completely new and untested program with no evidence of effectiveness, you may decide to aim for a lower evidence rating (e.g., “promising,” “supported”) due to the lower threshold of rigor—and therefore, lower investment of resources—required to achieve these ratings. However, if your program has already been designated as “promising” in the Prevention Services or CEBC clearinghouses, then it may be worth the investment of time and resources to seek a higher evidence rating by conducting an RCT (required to receive the CEBC’s highest rating) or a high-quality quasi-experimental design (which can support the Prevention Services Clearinghouse’s highest program and service rating).

Finally, it is also important to note that one study or program evaluation, even of the highest quality and using the most rigorous research design, does not by itself guarantee a clearinghouse review or a high evidence rating. For example, both the Prevention Services Clearinghouse and the CEBC require a program to have at least two documented studies that meet their respective eligibility criteria and evidentiary standards to be considered for their highest evidence ratings. As such, a child welfare intervention will ideally be implemented, studied, and documented in the literature multiple times to build a more conclusive record of effectiveness.

Next Steps

Once you select an appropriate research design, the work of planning and implementing a methodologically rigorous research study or program evaluation can begin. This requires the coordination and monitoring of multiple complex activities, such as ensuring the integrity of the random assignment process (in the case of RCTs), confirming the baseline equivalence of the intervention and control/comparison groups, mitigating selection bias, and maximizing enrollment in the program while minimizing attrition. These topics are beyond the scope of this brief; however, they are addressed in more detail in two JBA resources: [Conducting Randomized Controlled Trials in Child Welfare Practice Settings: Challenges and Solutions](#) and [Working With Small Samples](#). These publications, along with the other information and steps outlined in this brief, provide a starting place for the (at times) challenging but critical work of building the evidence base for effective programs and practices in the child welfare field.

References

- California Evidence-Based Clearinghouse for Child Welfare (2020). *CEBC ratings policy and procedures manual* (Rev. ed.). Chadwick Center for Children and Families, Rady Children's Hospital – San Diego.
- 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.
- Framework Workgroup. (2014). *A framework to design, test, spread, and sustain effective practice in child welfare*. Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services.
- James Bell Associates. (2013). *Conducting randomized controlled trials in child welfare practice settings*. Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services.
- James Bell Associates. (2018). *Formative evaluation toolkit: A step-by-step guide and resources for evaluating program implementation and early outcomes*. Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services.
- James Bell Associates. (2020). *Selecting and working with an evaluation partner*. Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services.
- James Bell Associates. (2013). *Working with small samples*. Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services.
- Wilson, S. J., Price, C. S., Kerns, S. E. U., Dastrup, S. D., & Brown, S. R. (2019). *Title IV-E prevention services clearinghouse handbook of standards and procedures, version 1.0*. OPRE Report # 2019-56. Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Submitted to

Tracy Y. Fields, M.S.W., M.A.
Child Welfare Program Specialist
Children's Bureau
Administration for Children and Families
U.S. Department of Health and Human Services
Contract Number: 47QRAA20D0008
Task Order Number: 140D0420F0597

Prepared by

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

Elliott Graham, Ph.D.
Project Director

This report is in the public domain. Permission to reproduce is not necessary. Suggested citation: Miller, N., & Graham, E. (2021). *Evidence-building evaluation strategies for untested and promising child welfare programs*. Children's Bureau, U.S. Department of Health and Human Services.

Disclaimer

The views expressed in this publication do not necessarily reflect the views or policies of the Children's Bureau, the Administration for Children and Families, or the U.S. Department of Health and Human Services. For more information, please visit the Children's Bureau website at <https://www.acf.hhs.gov/cb>.

