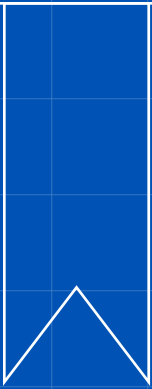**Chapter 1**

Defining Evidence of Effectiveness

Why This Matters

If building evidence of effectiveness were easy, this guide wouldn't exist. The reality is there are dozens of factors that contribute to evaluation capabilities and readiness, and the ability of organizations to make data- and evidence-driven decisions. Some are relatively straightforward, such as the ability to collect and use data, and budgeting for evaluations. Others may be less obvious but are equally important — such as defining what constitutes evidence of effectiveness.

Evidence and Evaluation Framework

To aid government organizations in their efforts to build evidence and evaluation capabilities and a culture supporting them, Results for America created an evidence and evaluation framework spanning four capacity areas: Foundations, Policies & Practices, Leadership & Culture and Evaluation Planning.

The strategies and skills within each component in the framework build upon and reinforce each other. Together, they give teams the ability to know what works best to equitably improve outcomes for residents and communities.

As you work to build your organization's evaluation capabilities and culture, use this framework — which is the product of conversations with jurisdictions around the country and a review of best practices — to guide your work. Consider using it to identify which skills are already present in your organization and which are not. Also consider conducting a needs assessment to support this process — it can be difficult to move forward if you don't know where you are.[1]

Subsequent chapters of this guide focus on specific elements of the framework, including:

- **defining equity (Chapter 2)**

- **collecting community input (Chapter 3)**

- **establishing and implementing an evaluation policy (Chapters 4 and 5)**

- **getting internal buy-in (Chapter 5)**

- **leveraging evaluation opportunities (Chapter 6)**

- **making evidence-based funding decisions (Chapter 7)**

- **building a culture of evidence and evaluation (Chapter 8)**

The rest of this chapter focuses on defining evidence of effectiveness, an important step toward establishing an evaluation policy.

Existing Definitions of Evidence

For policymakers and those who deliver services to have a shared understanding of which interventions are "evidence-based," they must have a shared definition of evidence of effectiveness — of what works.

As your organization works to create its own definition of evidence of effectiveness, consider adopting or adapting an existing framework. There is no reason to start from scratch — many federal agencies and state governments have already adopted definitions.

North Carolina's Office of State Budget and Management, for example, defines evidence as "the available body of facts or information indicating how likely it is that a belief is true. Evidence can be qualitative or quantitative, and it may come from a variety of sources, with varying degrees of credibility."

Read on for other examples.

Federal Examples

U.S. Department of Education
The Every Student Succeeds Act (ESSA) includes four levels of evidence, determined by study design, study results, negative findings from related studies, sample size and setting, and the match between study population and setting and the population and setting for implementation.

Evaluations: The Key to Understanding What Works

Evaluation: "An assessment using systematic data collection and analysis of one or more programs, policies, and organizations intended to assess their effectiveness and efficiency."[2]
—Office of Management and Budget

To build an understanding of which programs work for whom and under what circumstances, organizations must rigorously evaluate program effectiveness. Through a systematic analysis of high-quality quantitative or qualitative data, evaluations produce evidence of outcomes that is valuable in a number of important ways. Evidence of what works provides a foundational understanding of how to improve existing initiatives, enabling organizations to make evidence-based decisions that invest resources in proven programs.

More broadly, by building an understanding of what works and how, evaluations help to create equitable change at scale. It's important to note that there are multiple types of evaluations. To learn about impact evaluations and process evaluations, and how they differ, see Chapter 6.

U.S. Department of Treasury

Treasury's American Rescue Plan Reporting and Compliance Guidance defines three tiers of evidence: strong, moderate and preliminary. (See "Evidence Definition Spotlight" on p. 12 for details.)

AmeriCorps

The AmeriCorps State and National grant program uses four tiers of evidence (strong, moderate, preliminary and pre-preliminary evidence) based on study design, number of studies, and findings. It defines "evidence-based programs" that have "been rigorously evaluated and have demonstrated positive results for at least one key desired outcome. Rigorous evaluation means conducting at least one Randomized Controlled Trial (RCT) or Quasi-Experimental Design (QED) evaluation of the same intervention described in the application."

U.S. Department of Labor

The Clearinghouse for Labor Evaluation and Research rates evidence for interventions based on study quality, number of studies and breadth of favorable findings from the studies. Rating categories are high, moderate, potentially promising and no rating.

State Examples

State government actions to support definitions of evidence of effectiveness include:

Minnesota

The state assigns evidence ratings to programs, based on impact evaluations. Ratings include "proven effective" and "promising," among others. "Proven effective" is defined as offering "a high level of research on effectiveness for at least one outcome of interest. This is determined through multiple qualifying evaluations outside of Minnesota or one or more qualifying local evaluation."

Colorado

In 2021, the state's legislature passed a bipartisan bill requiring consistent definitions of evidence-based programs in budget requests. The state has created an "evidence continuum" to establish standards for building evidence to assess whether programs work.

Tennessee

The state maintains a comprehensive program inventory of state-funded programs, assigning each to an "evidence step." The highest step is "strong evidence."

New Mexico

A state law requires that agencies prioritize "evidence-based" programs that are "demonstrated to be effective for the intended populations through scientifically based research, including statistically controlled evaluations or randomized trials."
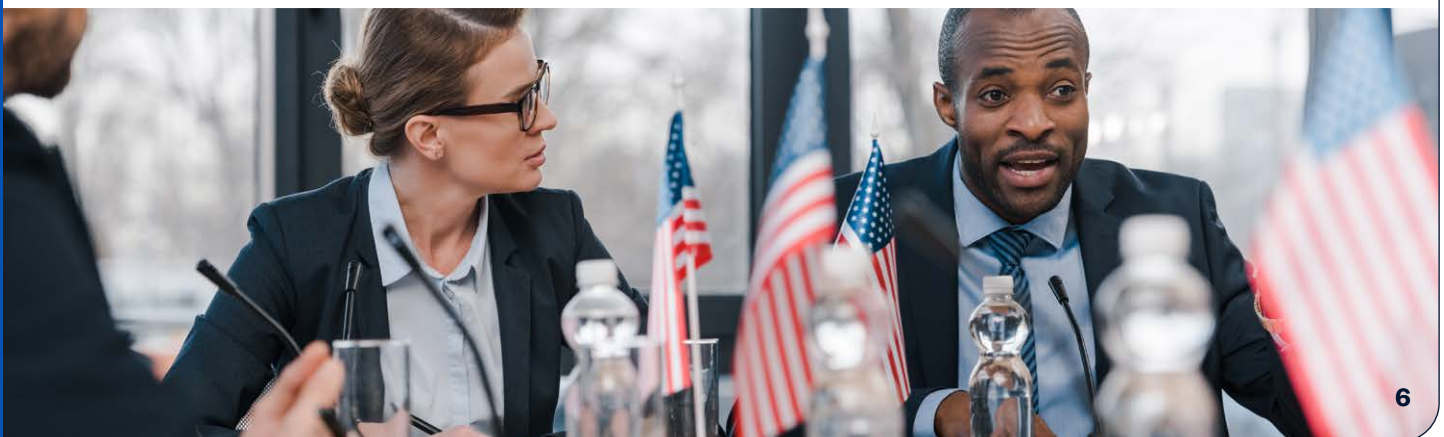
Evidence Definition Spotlight: U.S. Treasury Department

The American Rescue Plan Act established the Coronavirus State and Local Fiscal Recovery Funds ("SLFRF") program. As part of its reporting and compliance guidance for the program, the U.S. Treasury Department offered jurisdictions the following definitions of evidence, noting that to be designated "evidence-based," an intervention must have strong or moderate evidence.

- **Strong evidence** means that the evidence base can support causal conclusions for the specific program proposed by the applicant with the highest level of confidence. This consists of one or more well-designed and well-implemented experimental studies conducted on the proposed program with positive findings on one or more intended outcomes.

- **Moderate evidence** means that there is a reasonably developed evidence base that can support causal conclusions. The evidence base consists of one or more quasi-experimental studies with positive findings on one or more intended outcomes OR two or more non-experimental studies with positive findings on one or more intended outcomes. Examples of research that meet the standards include: well-designed and well-implemented quasi-experimental studies that compare outcomes between the group receiving the intervention and a matched comparison group (i.e., a similar population that does not receive the intervention).

- **Preliminary evidence** means that the evidence base can support conclusions about the program's contribution to observed outcomes. The evidence base consists of at least one non-experimental study. A study that demonstrates improvement in program beneficiaries over time on one or more intended outcomes OR an implementation (process evaluation) study used to learn about and improve program operations would constitute preliminary evidence. Examples of research that meet the standards include: (1) outcome studies that track program beneficiaries through a service pipeline and measure beneficiaries' responses at the end of the program; and (2) pre-and post-test research that determines whether beneficiaries have improved on an intended outcome.

Results for America's Definitions

Results for America's definitions of "evidence-based programs" and "evidence-building programs" are designed to help leaders steer taxpayer dollars to programs that will deliver better results for all. These definitions were developed in consultation with more than 80 stakeholders, including government officials, community advocates and practitioners, working in the field of evidence-based policymaking.

"Evidence-based program" means a program with either impact evidence or implementation evidence that is relevant and credible and has an informed rationale.

"Evidence-building program" means a program that has an informed rationale and is undergoing an impact evaluation or implementation evaluation that is relevant and credible.

"Impact evidence" means that the full body of evidence for a program shows that the program was very likely to have caused improvement on an important outcome in similar contexts and for similar populations, based on one of the following categories of evaluation findings:

- Category A: At least three well-designed and implemented quasi-experimental or experimental design studies from more than one site that show the program caused a statistically significant positive effect on an important outcome.

- Category B: One or two well-designed and implemented quasi-experimental or experimental design studies that show the program caused a statistically significant positive effect on an important outcome.

"Implementation evidence" means a program has one or more well-designed evaluations using quantitative, qualitative or mixed methods designs that indicate, in similar contexts and for similar populations, how well the program has been implemented, barriers that have been experienced during implementation, who the program has served, cost of implementing, who values the program, non-causal results associated with program implementation and/or other information that can be useful for program improvement and successful implementation in other settings.

"Informed rationale" means the reasoning (such as a theory of change, logic model or narrative description) behind why a program is likely to improve important outcomes in similar contexts and for similar populations, based on research and input from participants and relevant stakeholders.

Among other goals, these definitions are meant to elevate the importance of generating knowledge that focuses on why, how and for whom programs work, as well as encourage ongoing evaluation. Learn more here, including how Results for America's definitions can be used to accomplish different goals.

→ **For endnotes, see the full policy guide here.**

Exercise 1: Evidence Definition Workshop

To understand which policies and programs work and how, organizations need a shared understanding of what constitutes evidence of effectiveness. Even if your organization has adopted a definition of evidence, it is worth reviewing it periodically to ensure alignment with current best practices.

1. **Start by putting your organization's current definition of evidence on a whiteboard. If your organization doesn't have one, feel free to choose a definition adopted by another jurisdiction as a starting place. (See examples earlier in this chapter.)**

2. **Consider these questions, putting any thoughts and ideas in sticky notes below**

the definition:

- What do you like about the definition?

- What are its limitations? (For example, does it differentiate between tiers of evidence strength such as "strong" or "moderate," or categories of evidence such as "impact" or "implementation"?

- Does the definition make clear that evidence should be generated from formal evaluations of programs? Does it detail which types of evaluations are required or preferred?

3. **Draft a revised definition of evidence, based on comments and feedback gathered on the sticky notes. Circulate to relevant stakeholders in the organization to gather feedback.**