

Introduction

As a nation, we owe it to all citizens to invest our resources in the most effective solutions to the problems we face. This is particularly true when it comes to supporting vulnerable children, families, and communities—an area of both tremendous need and expense that is critical to the health and growth of our nation.

Unfortunately, we cannot say with confidence today that we are making the most of our spending. Less than 1 percent of federal government spending is backed by even the most basic evidence of impact.⁴ It may be that many government programs are working. We just don't know.

Investing in more effective solutions will not be easy. For starters, there is significant inertia in the way that resources are spent today. Making changes will require political will to shift the funding as well as new skills among those who implement the changes. Additionally, it is not easy to identify what the most effective solutions are. Relatively few interventions have been rigorously evaluated; even fewer are proven to have positive results. In fact, most evaluations show mixed results. Therefore, determining whether something works is not black and white and depends on how one plans to use it. Furthermore, the needs and populations in our country are constantly changing, so our solution set must continually evolve and consider local context.

A marketplace focused on learning and continuous improvement

We must embrace an approach that focuses on learning and continuous improvement and is based on meaningful data, quality evidence, and rigorous evaluation. A healthy market for evidence on effectiveness is critical for two primary reasons. First, it gives decision makers—whom we define as policy makers, funders, and practitioners—the information they need to select the appropriate solution for their circumstances. Second, it sheds light on areas where there is currently not enough evidence on effectiveness and thus more innovation or evaluation is needed.

The purpose of such a market is continuous improvement. It is *not* the separation of interventions into two neat categories of those that work and those that

⁴ Robert Balfanz, *The Power of a Penny: Building Knowledge to Invest in What Works in Education*, Invest in What Works Policy Series, Results for America, http://results4america.org/wp-content/uploads/2014/07/RFAPolicyBrief_k.pdf.

The purpose of a ‘what works’ market is continuous improvement. It is not the separation of interventions into those that work and those that don’t work.

don’t work. Our research shows that effectiveness is far more nuanced (including dimensions such as evaluation rigor, impact, and replicability) and constantly evolving, and that the required

level of effectiveness depends on the application. Placing too much emphasis on judging interventions could discourage innovation, as new interventions by definition do not have evidence. Judging interventions could also discourage the sharing of negative, mixed, or null results for fear of negative repercussions.

Spotlight: Moneyball for Government Part 3—Continuous Learning

Several leading organizations are using data, evidence, and evaluation to learn how to improve outcomes. Watch this video to see more about the work of three such organizations—City Year, Communities in Schools, and LIFT.

For more information: <http://moneyballforgov.com/moneyball-for-government-video-series/>

An approach that emphasizes learning and continuous improvement depends on many factors, most notably the ability to learn from past evaluations. But such evidence is not readily accessible today. Moreover, we don’t know enough about who is looking for information or what are they looking for. Are they finding what they need? From what sources? To begin answering these questions, and ultimately help increase the flow of information, Results for America, in consultation with The Pew-MacArthur Results First Initiative (Results First)⁵, commissioned The Bridgespan Group to study this “market” for evidence on effectiveness. We use the term **market** because the ultimate objective is to connect and align the creators of evidence (or supply—e.g., clearinghouses, evaluators, nonprofits) with the decision makers (or demand—e.g., policy makers, practitioners, private and public funders).

The following sections outline the findings of this study on the market for evidence on effectiveness. We describe the complexity of this market—a long and fragmented supply chain ([Key Finding #1](#)) and several types of users with different needs ([Key Finding #2](#)). While recognizing that the market is growing and improving, we then identify six gaps in the marketplace that prevent supply from effectively meeting the requirements of demand ([Key Finding #3](#)). Finally, we describe nine specific opportunities to address these gaps and strengthen the growing market for evidence on effectiveness ([Recommendations](#)).

⁵ The Pew-MacArthur Results First Initiative is a project of The Pew Charitable Trusts and the John D. and Catherine T. MacArthur Foundation. The opinions expressed in this report are those of the authors and do not necessarily reflect the views of the Pew-MacArthur Results First Initiative, The Pew Charitable Trusts, or the John D. and Catherine T. MacArthur Foundation.

Methodology

To better understand the market for evidence on effectiveness, we gathered input from sources (supply), users (demand), and experts familiar with both.

We conducted in-depth research in two domains—K-12 education and child welfare—to ensure we understood the full complexity of the demand side. We chose these domains because they are both critical areas for families and communities but also differ significantly with regards to who makes decisions, how funding flows, and the prevalence of effective interventions. We interviewed more than 20 education and child welfare decision makers with various levels of responsibility (state, county, district, and local/school). We also interviewed 10 officials in different government roles, such as the mayor’s office.

When examining the supply side, we broadened our research beyond child welfare and education to include all domains. We focused particularly on **clearinghouses**, which, for this report, we define as: information sources that aggregate, standardize, review, and rate the evidence base of interventions, acting as repositories that provide input into the decision-making process. However, we did not complete a broad scan of the full landscape of information sources in the United States. We also identified a subset of 14 clearinghouses for more in-depth research and interviews.⁶ This subset includes those that are larger and more well-known, but also covers a variety of domains and includes both public and private clearinghouses. These 14 will be referenced to a greater extent throughout the report. Finally, we conducted 11 additional interviews with leaders of other information sources, including two international clearinghouses, five purveyors (developers or providers of social programs or interventions), and four advisers and researchers.

While focused on supply and demand, we also interviewed 10 experts with broader perspectives of the market dynamics. In total, we completed more than 80 interviews; the full list of organizations interviewed is available in [Appendix 1](#).

Defining a common language

Our research revealed there is not a common language to discuss the topic of effectiveness. Terms like “evidence-based practices,” “evidence-based programs,” and “what works” are often used but seldom defined or differentiated. One purveyor noted, “If anything can help in this space, it would be adopting a national standard on what the term ‘evidence-based’ means.”

6 Blueprints for Healthy Youth Development; California Evidence-Based Clearinghouse for Child Welfare; Child Trends LINKS; Clearinghouse for Labor Evaluation and Research; Coalition for Evidence-Based Policy; CrimeSolutions.gov; FindYouthInfo.gov; Home Visiting Evidence of Effectiveness; National Registry of Evidence-based Programs and Practices; Office of Juvenile Justice and Delinquency Prevention’s Model Programs Guide; Promising Practices Network; Teen Pregnancy Prevention; Washington State Institute for Public Policy; and What Works Clearinghouse

There are two particular areas of inconsistency in discussing what works. This first is the definition of what is being evaluated. People refer to both “evidence-based practices” and “evidence-based programs.” Some use these terms interchangeably. Others give them different meanings, distinguishing *practices* as components of *programs*. Still others use the terms to mean other things. Given that both practices and programs can be evaluated, for this report we use the term **intervention** to refer generally to both practices and programs.

Additionally, there is confusion about the level of evidence required for an intervention to be deemed effective or something that works. First, there is not always a clear distinction between the *rigor* of the evaluations (i.e., how confident one can be in the results) and the size of the *impact* attributed to the intervention. We found that information sources had different standards for these dimensions, and some only looked at either *rigor* or *impact*.

Further compounding this confusion is the growing (and justified) importance attributed to the *replicability* of an intervention, or the ability of an intervention to be implemented with fidelity and demonstrate effectiveness more than once. Some are beginning to acknowledge the importance of replicability in rating interventions, but there are not yet consistent standards for integrating it.

We also found that some information sources make an overall judgment as to whether or not an intervention works. This can be confusing, as each source has its own standards for determining if something works, and thus a single intervention can receive conflicting categorizations. In addition, a clear distinction that something does or does not work can rarely be made. Most evaluations and interventions show mixed results and are highly dependent on the population and outcomes targeted, as well as contextual factors. Moreover, this type of stark distinction can be harmful if it discourages people from sharing negative, mixed, or null results. Similarly, it can discourage innovation, as all new interventions begin with no evidence. The right level and type of evidence needed ultimately depends on the application of the intervention. For example, those focused on continuous improvement must consider *all* evidence and results upon which to build.

A promising path out of this confusion may lie in the development of a common evidence framework. Such efforts are underway both within and outside of the government, including the recently revised *Education Department General Administrative Regulations (EDGAR)*; the recently announced *Common Guidelines for Education Research and Development*; the *Cross-Agency Platform for Judging Evidence*; a proposed *Framework for Continuous Evidence Building*; and the *Nesta Standards of Evidence* in the UK (see details in [Appendix 2](#)). In particular, EDGAR creates the opportunity to apply a tiered-evidence framework across all discretionary grant programs to reward evidence-based programs and produce rigorous evaluation of what works in education.

Given that so many efforts are underway, we do not select a specific framework for effectiveness for use in this report. Rather we use a broad definition of **evidence on effectiveness** when examining the market, which includes information about the rigor of evaluation, impact size, and replicability.