

Redefining Evidence in Education: A Comparison of Results for America's New Evidence Definitions & the Every Student Succeeds Act

May 2024

The U.S. Congress [defined “evidence-based”](#) for the first time in federal education law in December 2015 with the Every Student Succeeds Act (ESSA). This was an important moment: Clearly defining what it means for a program to be “evidence-based” is a crucial first step in embedding evidence into a decision-making framework.¹ In response, state education agencies (SEAs), local education agencies (LEAs), and other education leaders and institutions—including Results for America—set to work to understand the new definition and what it would mean for how states and school districts invest federal funds.

Results for America (RFA) has long supported leaders at all levels of government—including in our [education system](#)—to invest taxpayer dollars in programs with a track record of effectiveness. RFA has [helped translate](#) the ESSA definition for the field, [elevate opportunities](#) for leveraging it to improve outcomes, and [support SEAs](#) interested in shifting more funding toward evidence-based approaches that are more likely to deliver results. Through our work, particularly with the [Invest in What Works SEA Fellowship program](#) and the [Invest in What Works Coalition](#), RFA has had a firsthand view of the successes and challenges of the ESSA definition and its implementation.

Nearly a decade later, we can point to a [growing number of examples](#) of how ESSA's evidence definition is being used to produce positive [education outcomes](#). But, as is the case with most new federal policies, there have also been challenges. That's why, in 2021, we decided to revisit ESSA's definition to see if a revised approach could better address today's important questions.

In February 2024, we released [new definitions of “evidence-based” and “evidence-building”](#) for use across all policy areas and throughout all levels of government. (See the box on the next page for more on our process.) In capturing learnings from the last decade, we believe these new definitions will better support government leaders to learn, improve and invest in what works. This paper reviews RFA's new definitions and compares them to ESSA's definition.

¹ “Program” means an activity, strategy, intervention, practice or policy that has defined core features.

OUR PROCESS

When we [announced](#) our intention to revisit the existing definition of “evidence-based” in the education sector, we committed to conducting an inclusive and practical process. We spent months holding interviews and focus groups with educators, policymakers, researchers and community advocates and doing desk research. Midway through the process, we decided to expand our scope beyond education and the ESSA definition. Similar updates were needed in other sectors and aligning recommendations across policy areas would create beneficial consistency. Ultimately, we spoke with nearly 100 leaders from a broad range of backgrounds, and recommended that government leaders adopt new definitions of “evidence-based” and “evidence-building” that better support them to learn, improve and invest in what works.

WHY DO WE WANT TO REVISE ESSA’S DEFINITION?

ESSA’s definition has been catalytic in influencing how the federal government, state education agencies and school districts spend their limited resources. It also has clear limitations. As we note in our [2021 Moneyball for Education Policy Recommendations report](#), ESSA’s definition neither acknowledges the importance of considering a full range of evidence when selecting an intervention nor considers the relevance of the existing evidence to the challenge being addressed.

Challenges have also arisen from the field’s experience implementing the definition, whether in the context of SEAs writing consolidated plans, LEAs developing school improvement plans, or school leaders navigating the available evidence base to select particular programs. There are also important concerns about equity, recognizing that a wider array of methodologies capable of producing rigorous research findings can and should inform whether a program is working, as well as the need to understand why, how and for whom programs are working.

“

[ESSA’s definition] represented an important breakthrough in evidence-based policymaking, but it was only a first step that would need improvement, and there is a growing consensus that the time is ripe to improve it.”

—2021 Moneyball for Education Policy Recommendations

A revised definition could:

- Incorporate the full range and relevance of evidence.
- Align ESSA's programmatic definitions, the Institute of Education Sciences' research definitions and, ideally, the definitions in other federal statutes both within and beyond education laws.
- Make it easier for educators and policymakers to navigate the evidence base as they search for the best strategies to support students.
- Create space for the U.S. Department of Education and the broader education field to grapple with important questions about the relationship between evidence-based interventions and equity, including emerging questions about how to democratize evidence-based decision making.

While some of the limitations of ESSA's definition are fixable in the language of the definition, others are a result of the complexities of evidence-based policymaking, including limited capacity in the field for using evidence thoughtfully, shortcomings in the existing evidence base and systemic inequities. A definition of a key term—no matter how thoughtful—cannot fully address all of these challenges. But, we believe that updating the definition of evidence-based is an essential building block to continued progress.

NINE YEARS OF PROGRESS

ESSA's evidence definition and accompanying federal guidance have produced meaningful progress. At the state and local level this includes:

- Using the ESSA definition in school improvement plans and competitive grants to [shift more federal funds to evidence-based approaches](#).
- Using the tiered-evidence grantmaking structure in the federal [Education Innovation and Research](#) (EIR) program to direct larger grants to proposals with stronger supporting evidence.
- The creation and/or expansion of clearinghouses to [highlight evidence-based approaches](#).
- The integration of the definition into support structures and continuous improvement models.
- Embedding the definition of evidence into grants to school districts and in school district contracts.

NINE YEARS OF PROGRESS (CONTINUED)

RFA's [Honor Roll of State Grant Programs that Define and Prioritize Evidence](#) includes 17 examples of grant programs administered by SEAs that define and prioritize evidence of effectiveness as of October 2023. State departments of education have also formed communities of practice and peer networks, and created other opportunities to help educators and policymakers better understand and use evidence.

Meanwhile, all federal agencies, including the U.S. Department of Education (USED), are continuing to implement the [Foundations of Evidence-Based Policymaking Act](#). Recent federal actions that promote evidence-based policymaking in the education sector include:

- The requirement that SEAs use their 5% set aside from the American Rescue Plan Act's Elementary and Secondary School Emergency Relief Fund (ESSER) for evidence-based programs to address the academic impacts of lost learning time, and the requirement that LEAs use 20% of their funding for that purpose.
- USED's 2023 revision of its "[evidence-based](#)" guidance to further emphasize the importance of evidence in the use of federal education funds.
- The Biden Administration's emphasis on the centrality of investing in evidence-based approaches in its [Improving Student Achievement Agenda](#).

To date, RFA has identified 98 federal grant programs that define and prioritize evidence in grantmaking, including 20 programs, totaling \$3.6 billion, administered by USED.

DEFINING EVIDENCE-BASED & EVIDENCE-BUILDING

Our new definition is responsive to the many challenges we have observed in our work since ESSA's enactment and to the feedback we received in interviews and engagement with education leaders. One way it does this is by separately defining "evidence-based" and "evidence-building." Together, the two definitions are designed to steer public funding to programs and policies with a track record of success, while also encouraging leaders to:

- Encourage consideration of the **full body of evidence** for programs that are being considered.

- Give governments the option to encourage programs with a proven track record while recognizing and valuing the usefulness of a **broad set of methods** for understanding the successful implementation and potential impact of a given investment.
- Elevate the importance of generating knowledge that focuses on **why, how and for whom** programs work, given that an average effect can mask significant differences across subgroups, including by race, ethnicity, gender, geography, income and other characteristics.
- Distinguish evidence-based programs from programs that are not yet evidence-based, while creating **pathways for evidence-building**.
- Ensure the presence of an **informed rationale for implementation** in each context and based on stakeholder input. In other words, government leaders should be able to articulate a theory of change for why their proposed investment is likely to have a positive impact in their jurisdiction.
- Encourage **ongoing evaluation** of the implementation and impact of funded programs.



RESULTS FOR AMERICA'S NEW DEFINITIONS

“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.

“Impact evaluation” means an evaluation of effectiveness through a method that may result in the program meeting the impact evidence definition above.

“Implementation evaluation” means an effort to study how a program is being implemented and the conditions under which a program is likely to be successful.

“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.

“Program” means an activity, strategy, intervention, practice or policy that has defined core features.

We believe these new definitions will ultimately support more—and more effective—use of evidence-based approaches to improve outcomes for communities across the nation, especially for students, families, and educators working in our nation’s schools. Notably:

The definitions recognize that relevant, credible knowledge is generated by multiple methods.

The definitions clarify that there are different forms of evidence—impact evidence and implementation evidence—which help decision-makers understand both *whether* a program has been effective and *how* it can be implemented well. This is important because it recognizes that all rigorous research methodologies, including quantitative, qualitative and mixed methods approaches, can help inform government decisions in different ways, as long as they are well-designed and well-implemented.

The definitions emphasize the importance of matching the type of evidence with the specific question being asked. The definitions are designed to encourage policymakers to consider which type of evidence is best for their particular use case and context—including the population(s) being served—and match

this evidence to the specific issue or question a government or organization is seeking to address. This includes considering capacity to implement and the relationship between the proposed program and existing investments.

The definitions promote continuous learning and improvement by defining “evidence-building program” as a program with an informed rationale that is undergoing an evaluation, and by establishing clear categories for impact and implementation evaluation.

We defined “evidence-building program” so that leaders can design incentives and preferences for programs that are in the process of being evaluated (or, with the right support, could be on this path). Results for America strongly believes that governments should invest in testing innovative and evidence-based programs and building evidence about their effectiveness. In a policy area with a weak or no evidence base, government leaders may choose to conduct an implementation evaluation to inform improvements and ensure a program is being implemented well before conducting an impact evaluation to build new knowledge about whether the program produces a positive impact.

The definitions require there to be shared logic and supporting information for why a program will work in the proposed context. These two pieces are crucial to building confidence that a program is likely to work. If the intervention shows positive effects in one evaluation, it does not necessarily mean it will work in a different setting unless there is a theory-informed by research, participant experiences, and diverse stakeholder perspective—for why it will work in similar settings with similar populations.

The definitions are designed to be “shovel-ready” for use in a variety of ways to help leaders advance evidence-based policymaking across the government. We have provided [examples](#) for how these definitions can be used to simultaneously prioritize programs with impact evidence in budget decisions, highlight implementation evidence for service providers who will be implementing the funded programs, and award bonus points in grant applications for programs that have implementation evidence and are currently undergoing an impact evaluation, or are willing to commit to one as a condition of funding.



COMPARING THE NEW DEFINITIONS TO ESSA'S DEFINITION

The new definitions build on the strong foundation of ESSA and improve it in a number of important ways, including:

- All new categories require an informed rationale, which is not required in ESSA except for a similar requirement in Tier 4.
- All categories call for evidence that is [relevant](#) for the context and population/(s) that will be served.
- Impact evidence, Category A sets a higher bar than ESSA's Tier 1 because it requires three or more qualifying studies — in other words, it distinguishes programs with robust supporting evidence from programs with a smaller number of supporting studies. Additionally, both categories of impact evidence require consideration of the [full body of evidence](#) supporting a particular program.
- The implementation evidence definition embraces a broader array of research methodologies than might currently qualify as Tier 4 under ESSA.
- RFA's definition does not consider innovative approaches to be “evidence-based” in the way some innovations can qualify as Tier 4 under ESSA, though they may fall under our definition of “evidence-building.” This should not discourage innovation. Instead, it reflects RFA's strong belief that government should [invest funding in testing innovative](#) programs with an informed rationale and build evidence about their effectiveness.

The table on the next page details how RFA's new definitions relate to ESSA's current definition:

COMPARING THE NEW DEFINITIONS TO ESSA'S DEFINITION (CONTINUED)

ESSA's 4 Tiers of "Evidence-Based"	RFA's New Definitions	Notes
<p>Tier 1 ("strong") A statistically significant effect on improving student outcomes or other relevant outcomes based on at least 1 well-designed and well-implemented experimental study (i.e., randomized)</p>	<p>A program with Tier 1 evidence will be considered "evidence-based with impact evidence" if (1) the supporting "impact evidence" is relevant, including in similar contexts and for similar populations, (2) there is an "informed rationale" and (3) the full body of evidence supports the program's effectiveness.</p> <ul style="list-style-type: none"> ● If there are 3+ such studies from more than one site, then the program would be in Category A of "impact evidence." ● If there are 1-2 of these kinds of studies, then the program would be in Category B of "impact evidence." 	<p>Because our new definition of "impact evidence" treats experimental and quasi-experimental studies similarly, a program can add together both types of studies when qualifying for Category A or Category B. For example, a program supported by one experimental study and two quasi-experimental studies could qualify as Category A.</p>
<p>Tier 2 ("moderate") A statistically significant effect on improving student outcomes or other relevant outcomes based on at least 1 well-designed and well-implemented quasi-experimental study (i.e., matched)</p>	<p>A program with Tier 2 evidence will be considered "evidence-based with impact evidence" if (1) the supporting "impact evidence" is relevant, including in similar contexts and for similar populations, (2) there is an "informed rationale" and (3) the full body of evidence supports the program's effectiveness.</p> <ul style="list-style-type: none"> ● If there are 3+ such studies from more than one site, then the program would be in Category A of "impact evidence." ● If there are 1-2 of these kinds of studies, then the program would be in Category B of "impact evidence." 	
<p>Tier 3 ("promising") A statistically significant effect on improving student outcomes or other relevant outcomes based on at least 1 well-designed and well-implemented correlational study with statistical controls for selection bias</p>	<p>A program with Tier 3 evidence will be considered "evidence-based with implementation evidence" if (1) the supporting "implementation evidence" is relevant, including in similar contexts and for similar populations and (2) there is an "informed rationale."</p>	<p>Note that correlational studies qualify as "implementation evidence" supporting the program; both impact evidence and implementation evidence can lead to a program meeting the definition of "evidence-based."</p>
<p>Tier 4 To qualify as Tier 4, ESSA requires that an approach both:</p> <ul style="list-style-type: none"> ● Demonstrates rationale based on high-quality research or positive evaluation that such activity, strategy, or intervention is likely to improve student outcomes ● Includes ongoing efforts to examine the effects of such activity, strategy, or intervention 	<p>Programs with Tier 4 evidence will be considered "evidence-based with implementation evidence" if (1) the supporting "implementation evidence" is relevant, including in similar contexts and for similar populations and (2) there is an "informed rationale."</p> <p>If there isn't an existing evaluation that meets our "implementation evidence" definition, then a program that is currently considered to have Tier 4 evidence could instead be considered an "evidence-building program" if it also has an informed rationale and is undergoing an "impact evaluation" or "implementation evaluation."</p>	<p>Note that RFA's definition of "evidence-building programs" can be used to promote continuous learning and improvement across all programs, whereas this is only present in ESSA Tier 4.</p>

CONCLUSION

To accomplish our ambitious goals for [education and economic mobility](#), we cannot rely on any single policy –and certainly not on a single definition of a policy term. But words matter. RFA’s new definitions of “evidence-based” and “evidence-building” represent an evolution of evidence-based policy deeply informed by experience and input from a wide array of stakeholders. We hope they will help SEAs, LEAs, schools, educators and others involved in education invest in [evidence-based approaches](#) in a new, more nuanced, and more participatory way, taking care to both attend to prior research and to the voices of those closest to the challenges they seek to address.

We encourage you to explore these definitions and [consider the ways](#) results-driven governments can define and prioritize evidence to achieve better results for all. We also encourage you to reach out and share your experience implementing these definitions in your community. RFA intends for these definitions to be “living and breathing.” As we partner with government leaders to implement them, we will review and refine the definitions based on what we learn.

If you would like to learn more about how you can use these definitions in your jurisdiction, please [contact](#) Sara Kerr, Vice President, Education Policy Implementation at Results for America. We will support the education field in considering these new definitions and how they can best coexist with the existing ESSA definition of “evidence-based” –there’s no time like the present to strengthen our approach to making the best possible use of evidence to help each and every student thrive.