Designing Scalable and Generalizable Evaluations

ipa Innovations for Poverty Action

PRACTICAL GUIDE

DECEMBER 2025

Why This Guide

Impact evaluations have generated valuable evidence on what works to reduce poverty. Yet many programs that show strong results in small pilot studies lose their effectiveness once they are expanded to reach more people. This *scale-up effect*—a drop in impact when programs move from controlled pilots to large-scale implementation—has been documented across many areas, such as early childhood and education.[1]

One reason is that many evaluations are designed to understand whether a program works in a specific setting while overlooking how it might perform in new or larger contexts. As a result, even high-quality evidence can be difficult to apply when governments or organizations try to expand a program. For example, a pilot might work well when delivered by highly trained staff under close supervision but produce weaker results when implemented through existing public systems with limited resources.

Planning evaluations with scale in mind from the start can help avoid this problem. By considering how a program will operate under real-world conditions—such as in different or larger populations, or other delivery systems—evaluations can generate lessons that remain relevant when decision-makers consider expanding successful pilots.

Purpose

This guide helps researchers, implementing organizations, and funders design evaluations that better inform decisions about *scaling up*—that is, expanding a program to reach more people or new settings. It focuses on two key ideas:

- **Scalability:** whether a program can reach a large and growing number of people in need while sustaining its impact.
- **Generalizability:** whether findings from one place or group can be applied to others.

Both concepts relate to a broader goal: making evidence more useful for real-world decision-making.

Policymakers and funders often need to decide whether to expand a program, adapt it, or invest in something else. Evaluations that focus only on a pilot's narrow impact may show *what works* in one place, but not h*ow to make it work elsewhere* or at scale.

Integrating scalability and generalizability into evaluation design strengthens the extent to which results hold under typical, rather than ideal, conditions. This makes evidence more relevant for public policy, where programs must operate within budget limits, staff constraints, and political realities.

Design Choices That Strengthen Evidence For Scale

Each design decision—from how participants and sites are selected to how outcomes are measured—can improve or weaken the usefulness of evidence for scale. For example:

- **Sampling:** Selecting study locations and participants that closely match the characteristics of the population the program will eventually serve helps ensure that the evaluation results reflect the conditions likely to exist during large-scale implementation.
- **Program design:** Testing simplified versions or different delivery models can reveal what is more efficient.
- **Measurement:** Using tools that are feasible and affordable at scale enables continued learning and adaptation after the evaluation ends.

The guide encourages teams to think of evaluations as the first step toward scale, not as isolated studies. It highlights key design choices and trade-offs that can make results more useful for real-world decisions—in particular, identifying which parts of a program are essential, designing monitoring systems that can be maintained by governments, anticipating how costs might change as programs expand, planning for realistic implementation conditions and measuring indirect effects. By addressing these concerns early, teams can produce evidence that helps determine not only *what works*, but also *how to make it work at scale* and in *different contexts.*

About This Guide

The guide is organized in two main parts:

1. **Summary checklist** – a short overview of the main elements that affect whether evaluation results will be useful for scaling up. It can be used as a quick reference when planning or reviewing an evaluation.
2. **Detailed guidance** – sections that explain each element in more depth, including:
    - Identifying the **core components** of a program: the parts that are essential for success.
    - Designing **measures that can be scaled**: ways to track program quality that are practical in large systems.
    - Considering **economies and diseconomies of scale**: how costs change as a program expands.
    - Ensuring **representativeness**: making sure the study population and delivery conditions reflect those of real-world implementation.
    - Accounting for **spillover effects**: when the program indirectly benefits or harms people who are not direct participants.

Each section includes guiding questions, examples, and practical advice from real evaluations that have addressed these challenges.

The guide is grounded in the growing literature on the "science of using science." Our overall framing of threats to scalability and generalizability, draws heavily on Al-Ubaydli, Lee, List, and Suskind (2021)[2] and List (2022)[3]. For each design dimension, we build in particular on: (1) the notion of core components and minimum effective packages developed in this literature and in Fixen et al. (2005)[4]; (2) Caron, Bernard, and Metz (2021)[5] on measuring fidelity with tools that are feasible to sustain at scale; (3) Davis et al. (2021)[6] on economies and diseconomies of scale; (4) Davis et al. (2021)[7] and Caron, Bernard, and Metz (2021)[8] on the representativeness of the study population and implementation conditions ("properties of the situation"); and (5) Momeni and Tannenbaum (2021)[9] on accounting for spillover effects when evaluating programs at scale.

Our contribution is to translate these insights into a practical checklist and concrete guidance for teams planning impact evaluations with scale in mind.

Summary Checklist

The checklist below summarizes the main design elements that influence whether an evaluation can produce evidence that is useful for scaling up. Each element includes a guiding question, why it matters, and practical direction for teams planning or reviewing an evaluation.

Element	Key Question	Why It Matters	Practical Guidance
1. Core components	Have the program's essential features been clearly identified?	Knowing which parts of a program are essential helps preserve its impact when it expands.	Work with implementers and researchers to identify and test which components are non-negotiable for success. Simplify where possible to create a "minimum effective package."
2. Measures that can be scaled	Are the tools for tracking program quality realistic to use at scale?	Monitoring systems used in pilots may be too costly or complex for large-scale delivery.	Design fidelity measures that rely on simple, reliable data sources such as administrative records or brief observation tools. Validate them early.
3. Economies and diseconomies of scale	Is the intervention designed to maintain —or even improve—its cost-effectiveness as it scales?	Costs and efficiencies often change as programs expand. Ignoring these dynamics can lead to under- or over-estimating real-world feasibility.	Identify which costs will fall (e.g., materials, technology) and which may rise (e.g., supervision, skilled staff). Consider how technology, training, or simplified delivery models can offset higher costs.
4. Representativeness of the population	Do the study sites and participants reflect those the program aims to reach at scale?	Results from a non-representative pilot may not apply when a program reaches new or broader groups.	Select study areas and participants that share key characteristics—such as geography, demographics, or service access—with the intended scale-up population. Use administrative data to guide this selection when possible.
5. Representativeness of the situation	Will the delivery conditions during the pilot resemble those under real implementation?	Pilots often operate under ideal conditions that differ from government or routine delivery.	Whenever possible, use existing systems, staff, and infrastructure for implementation. Document differences between pilot and expected scale-up conditions.
6. Spillover effects	Could the program indirectly affect people who are not direct participants?	Ignoring spillovers can lead to over- or under-estimating total impact at scale.	Consider whether interactions between participants and non-participants might change outcomes. If likely, include plans to measure or account for these effects.

Designing for Scale: Key Considerations

Designing an evaluation that informs decisions about scale requires deliberate choices at every stage. The following areas—each with a guiding question, rationale, and practical steps—highlight where thoughtful design can make evaluation findings more relevant for real-world implementation.

1. Identify What Makes the Program Work

KEY QUESTION

Have the program's essential features been clearly identified?

WHY IT MATTERS

Programs often include multiple activities, but only some are critical for achieving impact. These core components are the mechanisms that must remain unchanged for the program to work.[10] If they are not identified early, scale-up efforts face two risks: removing or altering elements that are essential for impact, and including components that are costly or complex but not actually necessary, which reduces the likelihood of successful scale-up.

PRACTICAL GUIDANCE

- Work with implementers and researchers to map all components and determine which are essential.
- If it is unclear what drives results, design the evaluation to test different combinations of activities to find the "minimum effective package."
- Choose an approach that matches your resources and purpose. There are several ways to identify which components are essential:
    - **Review existing evidence** to develop hypotheses about what is core.[11]
    - **Test simplified versions of the program** to see whether results hold when certain elements are removed or adjusted.
    - **Compare different combinations of activities** in small pilots to understand which features truly drive results.
- Whichever approach is used, document how these decisions were made and note any limitations, so policymakers can judge how strong the evidence is for the final "minimum effective package."
- Once confirmed, monitor these core components consistently.

EXAMPLES

- BRAC's *Graduation programs* identified three essential elements: (1) a productive asset or cash grant, (2) temporary support for basic needs, and (3) regular coaching —from looking across a wide range of Graduation program variants that bundled different components (including others such as business capital, financial services, market linkages, and skills training) in different combinations. Streamlining to these non-negotiables helped preserve impacts while enabling the model to scale feasibly.[12]
- In Liberia's *School-Based Agricultural Extension* program, impacts on student learning only appeared when parental engagement was added to the other three components of the program: classroom extension, school demonstration farms, and student home projects., showing that what seemed optional was actually critical.[13]

2. Build Scalable Monitoring Systems

KEY QUESTION

Are the tools for tracking program quality realistic to use at scale?

WHY IT MATTERS

Pilots often rely on intensive oversight—frequent visits, detailed surveys, or external staff—that is not sustainable at scale. Yet maintaining impact at scale depends on implementation fidelity, which makes it essential to keep measuring quality as programs grow. Designing scalable measures means creating simple, reliable ways to track quality using the same systems that governments or organizations will rely on later.[14]

PRACTICAL GUIDANCE

- Use existing administrative data (e.g. routine records collected by schools or clinics) whenever possible and validate that data early on to ensure its accuracy and reliability and assess whether that data will continue to be collected and updated over time.
- In programs that require qualitative monitoring—such as early childhood or education—, keep observational tools short, and easy for regular staff to use, but also valid.[15] To reduce bias, steps may be needed when supervisors have close ties to those being observed—for example, rotating raters or triangulating with other data sources.[16]
- Emerging technologies, such as AI-assisted video analysis, also offer opportunities to automate supervision.
- Embedding behavioral incentives can also help maintain fidelity—through social comparisons, self-reporting tools, random checks or non-monetary recognition.

EXAMPLES

- In Ghana's Differentiated Learning program, monitoring tools developed for an earlier evaluation were simplified and adopted by the government's district education offices as part of regular supervision, enabling fidelity to the program to remain sustainable.
- In the Chicago Heights Early Childhood Center study, researchers led by John List tested how behavioral incentives could strengthen implementation at scale. Teachers were given a bonus upfront and told they would keep it only if their students met learning targets. The "loss-framed" incentive improved teacher effort and student learning compared to traditional end-of-year rewards. This illustrates how simple behavioral design—rather than heavy oversight—can sustain performance and fidelity as programs expand.[17]

3. Plan for Cost and Delivery at Scale

KEY QUESTION

Is the intervention designed to maintain—or even improve—its cost-effectiveness as it scales?

WHY IT MATTERS

When programs grow, some costs decrease (*economies of scale*) while others increase (*diseconomies of scale*). Understanding these patterns helps planners forecast realistic budgets and maintain quality as coverage expands.

PRACTICAL GUIDANCE

- Break down total costs into major components (e.g., design, materials, training, frontline delivery, supervision, technology and data systems, transport) and note which are mostly fixed and which are variable.
- For each component, consider how its cost per participant is likely to change as coverage expands (for example, design and platforms may become cheaper per person, while supervision or reaching harder-to-access communities may become more expensive).
- Explore the use of technology and simplified delivery, including services being effectively delivered by lower- or average-skilled workers, to keep per-person costs stable or declining.[18]
- Pay special attention to components that depend on scarce, highly skilled staff, intensive oversight or coordination, as these are particularly likely to drive up costs at scale and to create implementation challenges.
- Evaluate efficiency together with quality; cost reductions that undermine performance can erode impact.

EXAMPLES

- *Edutainment* programs in Tanzania and Uganda achieved large-scale reach at low cost by using radio and TV to deliver messages about gender equity.[19]
- Extensive evidence from *mental health and psychosocial support interventions* demonstrates that trained lay community members can effectively deliver services in low-resource settings, helping programs avoid diseconomies of scale.[20] This has been documented across multiple contexts, including cognitive behavioral therapy programs in Liberia showing that community workers, rather than professional psychologists, could effectively deliver sessions.[21]

4. Select a Realistic Study Population

KEY QUESTION

Do the study sites and participants reflect those the program aims to reach at scale?

WHY IT MATTERS

If the evaluation focuses on participants who are not representative of the population the program aims to serve, its results may not translate when scaled.[22]

PRACTICAL GUIDANCE

- Define the intended population for scale-up clearly—such as all public schools or low-income households in rural areas.
- Ideally, randomly select study sites and participants from within the target scale-up population.[23]
- When random selection is not feasible—for example, because some sites are already implementing the intervention, decline to participate, or because implementation would become too costly or logistically complex with a more dispersed sample—select sites and participants that share key characteristics with the target population, and use existing data to verify that alignment.[24]
- If the study sample differs from the scale-up population, document these differences so policymakers can interpret the findings accurately.

EXAMPLE

- In one African country's initiative to embed Teaching at the Right Level principles into its national remediation model, budget and operational constraints meant the pilot could run in only one district. Using national administrative data, IPA and the Ministry selected education zones whose schools closely matched the national average and mirrored the zone-level management structure that would oversee scale-up, making the pilot more informative for national planning. The impact evaluation planned for this project will apply the same logic to choose intervention and comparison schools while balancing restrictions with statistical power.

5. Test Under Real-World Conditions

KEY QUESTION

Will the delivery conditions during the pilot resemble those under real implementation?

WHY IT MATTERS

For results to meaningfully inform scale-up, the pilot must be delivered under conditions that the government or implementing agency can realistically sustain. Pilots often operate with more staff, more funding, or closer oversight than what is feasible in routine systems. However, the goal should be to make the evaluation as pragmatic as possible.[25] If the intervention is not designed to match the implementer's actual capacity—or if there is no credible plan to expand that capacity—pilot results may overstate what is achievable at scale and lead to misguided policy decisions.

PRACTICAL GUIDANCE

- Deliver the program through existing systems (e.g., government agencies or local organizations) whenever possible, rather than creating separate structures. This includes relying on existing infrastructure, staffing, and monitoring systems.
- If hiring implementation staff specifically for the evaluation, ensure their qualifications and time allocations match what would be feasible at scale, so that delivery and supervision conditions realistically mirror scale-up. If hiring implementation staff specifically for the evaluation, ensure their qualifications and time allocations match what would be feasible at scale, so that delivery and supervision conditions realistically mirror scale-up. One practical approach is to first define how staff would be recruited and selected at scale, identify the full pool of acceptable candidates under that process, and then randomly select from that pool to staff the evaluation, rather than recruiting a small team for the evaluation from the strongest candidates that that would be disproportionately better relative to scale conditions.[26]
- If the program plans to select providers, randomly select them to capture variation in delivery.
- Note any differences between the pilot and likely large-scale implementation—such as training intensity or incentives—and factor them into analysis and interpretation.
- The PRECIS framework (Pragmatic-Explanatory Continuum Indicator Summary) can help teams assess how closely evaluation conditions match real-world delivery systems.

EXAMPLE

- In Ghana's *Teacher Community Assistant Initiative* (TCAI)[27], using community volunteers to provide remedial tutoring to students produced strong results but was difficult for the government to sustain because of hiring costs. The teacher-led model was more feasible for national adoption, and a later evaluation[28] tested ways for school managers to better support teachers to deliver the model more effectively. This shows that, in retrospect, evaluations may benefit from focusing earlier on models that are more feasible to implement at scale.

6. Consider Spillover Effects

KEY QUESTION

Could the program indirectly affect people who are not direct participants?

WHY IT MATTERS

Programs can create *spillover effects*—indirect changes for non-participants. These can be positive (for example, siblings or friends of children enrolled in an early childhood education program also learn new skills) or negative (for example,the displacement of non-participants in a job creation program). Measuring them helps estimate the program's true total impact and prevents misleading conclusions as it expands.

PRACTICAL GUIDANCE

- Map likely spillover channels and anticipated magnitude early. Start by analyzing how participants might influence non-participants through social networks, shared markets, or public services. For example, students may influence classmates in an education program, beneficiaries may affect local prices in a livelihood intervention, or increased clinic use may occur in a conditional cash transfer program. Then, use theory and implementer experience to judge whether spillovers are likely to be meaningful.

- If spillovers are expected to be large, plan from the start to capture those effects. Momemi and Tannenbaum described three approaches researchers use to measure spillovers, which are described below. [29]

    - If spillovers could confuse results and are not the main focus of the study, design the program so that participants and non-participants have limited contact with each other—for instance, experiments involving resumes, letters or other nudge messages are well suited to minimize spillovers . This clarifies the program's direct effect, though it may reduce relevance for scale-up where such interactions are unavoidable.
    - When measuring spillovers separately is infeasible but they are expected to occur, randomize at a higher administrative or social level (e.g., school, village, or market). This captures both direct and indirect effects together under the assumption that most spillovers occur within, not across, those units.
    - If understanding spillover dynamics is essential and resources allow, use a two-stage design: Stage 1: Randomly assign clusters (e.g., villages or schools) to treatment or comparison. Stage 2: Within treated clusters, randomly assign only some eligible individuals to participate. This enables separate estimation of direct and spillover effects but requires larger samples and higher costs.

- Choose the approach based on expected spillover magnitude and available resources. Designs that control for spillovers improve precision, while those that measure them enhance policy relevance for scale-up.

EXAMPLE

- In a *multi-country study of the Graduation approach*[30], researchers measured possible effects on non-participants—like changes in wages or livestock prices—and found no negative impacts. Future analyses found that targeting villages with the highest poverty rates was actually more cost-effective than individual households. This helped confirm that the model could scale safely.

Conclusion and Key Takeaways

Impact evaluations are powerful tools for learning what works—but to inform real policy and large-scale delivery, they must also help us understand *how* and *under which conditions* programs can work at scale. Designing with scalability and generalizability in mind ensures that the lessons from pilots do not remain on the shelf but instead translate into guidance for governments and organizations seeking to expand effective solutions.

The principles outlined in this guide show that designing for scale is not a separate step—it is part of good evaluation practice. Each design decision, from defining core components to choosing study sites and measurement tools, shapes whether evidence will remain useful once programs move beyond controlled settings. Applying these principles leads to evaluations that bridge the gap between research and implementation, making evidence more actionable for policy.

Key Takeaways

- **Plan for scale from the beginning.** Treat impact evaluations as the first step toward expansion, not as stand-alone studies. Early design choices determine how relevant results will be for policy and real-world delivery.

- **Clarify what drives success.** Identify and protect the core components that are essential to achieving results, ensuring they remain consistent as the program grows.

- **Use systems that can be sustained.** Design monitoring and delivery structures that reflect real conditions, using existing staff, data systems, and resources.

- **Reflect real contexts and populations.** Choose study participants and settings that resemble those the program will reach at scale, so findings remain credible and applicable.

- **Consider broader effects.** Account for how scaling may change costs, and effects—both positive and negative—on people beyond direct participants.

Evaluations that anticipate scale do more than test whether a program works—they help decision-makers understand how to deliver results widely and sustainably.

References

1. Al-Ubaydli, O., Lee, S., List, J. A., & Suskind, D. L. (2021). The science of using science: A new framework for understanding the threats to scaling evidence-based policies. In J. A. List, D. L. Suskind, & L. H. Supplee (Eds.), The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it.
2. Ibid
3. List, J. A. (2022). The voltage effect: How to make good ideas great and great ideas scale. Currency.
4. Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation Research: A Synthesis of the Literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).
5. Caron, B., Bernard, S., & Metz, A. (2021). Fidelity and properties of the situation: Challenges and recommendations. In J. A. List, D. L. Suskind, & L. H. Supplee (Eds.), The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it. Routledge.
6. Davis, J., et al. (2021). Studying properties of the population: Designing studies that mirror real-world scenarios. In J. A. List, D. L. Suskind, & L. H. Supplee (Eds.), The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it. Routledge.
7. ibid
8. Caron, B., Bernard, S., & Metz, A. (2021). Fidelity and properties of the situation: Challenges and recommendations. In J. A. List, D. L. Suskind, & L. H. Supplee (Eds.), The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it. Routledge.
9. Momeni, F., & Tannenbaum, D. (2021). Spillovers and Program Evaluation at Scale. In J. A. List, D. L. Suskind, & L. H. Supplee (Eds.), The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it. Routledge.
10. Al-Ubaydli, O., Lee, S., List, J. A., & Suskind, D. L. (2021). The science of using science: A new framework for understanding the threats to scaling evidence-based policies. In J. A. List, D. L. Suskind, & L. H. Supplee (Eds.), The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it. Routledge.
11. Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation Research: A Synthesis of the Literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).
12. McAnnally-Linz, H. (2025). Introducing the Graduation essentials: Core components designed for scale. BRAC Ultra-Poor Graduation Initiative. Retrieved from https://www.bracusa.org/news/introducing-the-graduation-essentials-core-components-designed-for-scale /
13. Innovations for Poverty Action (IPA). (2024). Can school-based agricultural extension programs improve technology diffusion and rural education? Retrieved from https://poverty-action.org/can-school-based-agricultural-extension-programs-improve-technology-diffusion-and-rural-education
14. Caron, B., Bernard, S., & Metz, A. (2021). Fidelity and properties of the situation: Challenges and recommendations. In J. A. List, D. L. Suskind, & L. H. Supplee (Eds.), The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it. Routledge.
15. ibid
16. ibid
17. List, J. A. (2022). The voltage effect: How to make good ideas great and great ideas scale. Currency.
18. Davis, J., et al. (2021). Studying properties of the population: Designing studies that mirror real world scenarios. In J. A. List, D. L. Suskind, & L. H. Supplee (Eds.), The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it. Routledge.
19. Innovations for Poverty Action (IPA). (2021). How edutainment is a powerful tool to fight gender inequality. Retrieved from https://poverty-action.org/how-edutainment-powerful-tool-fight-gender-equality
20. Innovations for Poverty Action (IPA). (2023). Mental health and psychosocial support interventions in displacement settings: A scoping review. Retrieved from https://poverty-action.org/sites/default/files/2023-11/Displacement-Scoping-Review-November-2023.pdf
21. Blattman, Christopher, Sebastian Chaskel, Julian C. Jamison, and Margaret Sheridan. "Cognitive behavioral therapy reduces crime and violence over ten years: Experimental evidence." American Economic Review: Insights 5, no. 4 (2023): 527-545.
22. Al-Ubaydli, O., Lee, S., List, J. A., & Suskind, D. L. (2021). The science of using science: A new framework for understanding the threats to scaling evidence-based policies. In J. A. List, D. L. Suskind, & L. H. Supplee (Eds.), The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it. Routledge.
23. Davis, J., et al. (2021). Studying properties of the population: Designing studies that mirror real world scenarios. In J. A. List, D. L. Suskind, & L. H. Supplee (Eds.), The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it. Routledge.
24. ibid
25. Davis, J., et al. (2021). Studying properties of the population: Designing studies that mirror real world scenarios. In J. A. List, D. L. Suskind, & L. H. Supplee (Eds.), The scale-up effect in early childhood and public policy: Why interventions lose impact at scale and what we can do about it. Routledge.
26. ibid
27. Duflo, Annie, Jessica Kiessel, and Adrienne M. Lucas. "Experimental evidence on four policies to increase learning at scale." The Economic Journal 134, no. 661 (2024): 1985-2008.
28. Beg, Sabrin A., Anne E. Fitzpatrick, and Adrienne Lucas. Managing to learn. No. w31757. National Bureau of Economic Research, 2023.
29. Momeni, Fatemeh and Daniel Tannenbaum. “Spillovers and Program Evaluation at Scale.” In John A. List, Dana Suskind, and Lauren H. Supplee (eds.), The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It. Routledge, 2021.
30. Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., & Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. Science, 348(6236), 1260799.

Innovations for Poverty Action (IPA) is a research and policy nonprofit that discovers and promotes effective solutions to global poverty problems. IPA designs, rigorously evaluates, and refines these solutions and their applications together with researchers and local decision-makers, ensuring that evidence is used to improve the lives of the world’s poor. Our well-established partnerships in the countries where we work, and a strong understanding of local contexts, enable us to conduct high-quality research. This research has informed hundreds of successful programs that now impact millions of individuals worldwide.

Writers: Mercedes Sidders, Karla Petersen, Ana Tamayo

Editors: Ana Tamayo, Michael Podesta

Design: Ana Tamayo