Forging Collaborations for Scale: Catalyzing Partnerships Among Policy Makers, Practitioners, Researchers, Funders, and Evidence-to-Policy Organizations

Samantha Carter, Iqbal Dhaliwal, Samantha Friedlander, and Claire Walsh

Abdul Latif Jameel Poverty Action Lab (J-PAL), Department of Economics,

Massachusetts Institute of Technology (MIT)

Author Note

Samantha Carter is now at Founders Pledge.

We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Samantha Friedlander, 400 Main Street, E19-201, Cambridge, MA 02142. Email: sfriedlander@povertyactionlab.org

Abstract

Despite growing interest in evidence-informed policy making, understanding of how to scale up effective programs is still limited. Successfully scaling up an evidence-informed program or intervention requires: a deep understanding of both global evidence and the specific local context and systems, a policy window where change is possible, political will to change the status quo, adequate funding resources, and capacity to monitor and implement the program well. Because no single organization's mandate covers all of these conditions, collaborations between policy makers, researchers, practitioners, funders, and evidence-to-policy organizations can ensure that these scale-up conditions are met. In this chapter, we draw on the Abdul Latif Jameel Poverty Action Lab's (J-PAL) many years of experience in building scale-up collaborations to share key lessons we have learned about how to build scaling collaborations that combine the complementary skills and expertise each partner brings. The four key lessons are to: (a) invest in long-term partnerships and develop the resources to respond to policy windows in real time; (b) use several complementary types of data and evidence to inform every scaling effort; (c) help institutionalize a broader culture of evidence-informed policy making that goes beyond individual programs; and (d) leverage evidence-to-policy organizations that can play a critical role in bringing different stakeholders together to make change happen.

Keywords: randomized evaluation, evidence-based policy, scale-up

Introduction

Many chapters in this book focus on the technical process of scaling, including generating actionable evidence on the effectiveness of a program and ensuring that it is representative of the population and situation to which it will be scaled. This process is at the heart of what we do at the Abdul Latif Jameel Poverty Action Lab (J-PAL)—it is fundamental to the randomized evaluations conducted by our 400+ affiliated researchers and our network of research staff worldwide. Yet, the establishment of J-PAL and the focus of our work beyond research is an explicit recognition that scaling a program is also a political process that involves building relationships and partnerships with people in government and civil society, connecting individuals and organizations with different interests, persuading relevant decision makers, building a constituency for change, and mobilizing resources (Cooley et al., 2016). This work requires a complementary but different set of skills and incentives than researchers generating evidence need. In this chapter, we focus primarily on the political, human, and organizational elements of scaling that are sufficient conditions for success or failure after the necessary technical aspects of scaling have been met.

We define scaling broadly as expanding a program or policy to reach more people or changing an existing program, policy, system, or delivery model to make it more effective or allow it to reach more people. Our definition expands on that used by Al-Ubaydli et al. (2020) in Chapter 6, which focuses on the process of a government adopting a new intervention that has first been evaluated at a smaller scale, to include the process of changing an existing large-scale program to improve its effectiveness. Scaling can take place in the same or different context in which the program was originally developed and evaluated. We primarily focus on collaborations to scale up evidence-informed programs in partnership with government, since in most (if not all) countries, governments continue to be the biggest implementers of social policies and programs.

With offices at MIT and six other universities around the world, J-PAL is a global research center working to reduce poverty by ensuring that policy is informed by scientific evidence. Anchored by a network of 400+ affiliated professors at universities around the world, J-PAL conducts randomized impact evaluations to answer critical questions in the fight against poverty. We work directly with decision makers to translate this research into action, promoting a culture of evidence-informed policy making worldwide. Our policy analysis and outreach help governments, NGOs, donors, and the private sector apply evidence from randomized evaluations to their work and contributes to public discourse around some of the most pressing questions in social policy and international development. Finally, J-PAL works to build the capacity of researchers, policy makers and donors, and advocates of evidence-informed policy. We create university-level online courses and deliver in-person trainings to help people become better producers and users of evidence. Two of J-PAL's many initiatives that fund fieldwork are explicitly focused on scaling (our Innovation in Government Initiative [IGI] and Innovations in Data and Experiments for Action Initiative [IDEA]), which provide us with opportunities to learn about this challenging process.

In this chapter, we draw on more than a decade of J-PAL's experience worldwide to share key lessons we have learned about how to build scaling collaborations that combine the complementary skills and expertise each partner brings. After outlining the incentives and constraints of each type of actor, we discuss four broad lessons for forging successful scaling collaborations.

The four key lessons are to: (a) invest in long-term partnerships and develop the resources to respond to policy windows in real time; (b) use several complementary types of data and evidence to inform every scaling effort; (c) help institutionalize a broader culture of evidence-informed policy making that goes beyond individual programs; and (d) leverage evidence-to-policy organizations that can play a critical role in bringing different stakeholders together to make change happen. Some of these lessons seem obvious, but we continue to be surprised by the fact that they are overlooked in many scale-up efforts, when they could have prevented many subsequent failures.

The partnerships from which we draw these lessons span many years and continents, but they all demonstrate the importance of collaboration.[i] In each example, coalitions of governments, researchers, NGOs, funders, and evidence-to-policy organizations together navigated the complex processes of creating large-scale change in systems that otherwise preserve the status quo.

Why is collaboration critical to the success of taking evidence-informed programs to scale? Different stages of the scaling process require different skills, expertise, and resources. While some organizations can design evidence-informed innovations and pilot, test, and scale effective programs well on their own, most organizations do not have the mandate, incentives, skills, or resources to manage the entire scaling lifecycle. Meanwhile, policy makers frequently look to partner with civil society and research organizations outside government to help them solve important policy challenges.

Ultimately, scaling is both a science and an art. J-PAL has devoted considerable resources and efforts over the past two decades to the science part. However, our task in this chapter is to focus on the art of scaling, and how it can help us get the science of scaling right by bringing together the complementary contributions of a diverse set of actors.

Defining the Strengths, Incentives, and Constraints of Each Collaborator

In order to facilitate productive collaborations that lead to scale, all stakeholders must understand each other's perspectives. Who are the other partners? What will they contribute? What incentives drive them, and what constraints hold them back? We explore those questions below. The factors identified are meant to be a starting point, not an exhaustive list. In addition, many organizations fall into multiple categories. For example, governments often provide funding for implementation and evaluation by third-party organizations. Bilateral and multilateral organizations, including the Inter-American Development Bank, UK Department for International Development, United States Agency for International Development, World Bank, and others, carry out a broad range of functions including funding, implementation, and research, and thus do not fit neatly into a single category.

Governments

Governments manage vast bureaucracies and have deep experience implementing programs at scale. Their understanding of local context makes them well placed to contribute to the adaptation of evidence to local conditions and tailor programs to their systems, although they may not always have the time or resources to do so. Governments also bring the credibility necessary to bring together stakeholders. They write wide-reaching policies and programs, so having a government as a partner often means the potential to reach many more people than otherwise possible. Governments have an explicit mandate to improve the well-being of their citizens and are therefore almost always the biggest funder of social programs.

The incentives of policy makers in government are typically to show progress toward the government's goals, including improving people's lives. In addition to budget constraints that affect all organizations, policy makers have to operate under a unique set of constraints driven by politics (stated ideology or manifestoes), time (short deadlines and frequent election cycles), bureaucracy (procurement regulations), capacity (career civil servants with skill sets that may not be updated to the challenges), turnover (frequent transfers), and political sensitivity (media and opposition criticism).

Governments are not monoliths, and the factors described above may be different for different individuals or departments within the same government. Building coalitions for scale with a government frequently involves building partnerships with multiple departments, ministries, and individual officials at various levels, all of whom may have different interests and incentives.

Non-Profits or NGO Practitioners

Non-governmental organizations (NGOs) play an important role in scale-ups by scaling programs either themselves or in partnership with governments. NGOs' comparative advantages include implementation experience and expertise, local context and program-related expertise, commitment to mission, flexibility, program management capabilities, and a more direct connection with beneficiaries than any other partner in the scale-up collaboration.

The incentives of NGOs include showing progress toward their mission, improving people's lives, maintaining relationships with beneficiaries, and attracting and maintaining donor relationships.

Accordingly, the main scaling constraints of NGOs are typically related to their (or their founder's) ideology, fundraising, capacity to operate at a state or national scale, and challenges accessing governments' existing infrastructure (for example, public schools or hospitals) in which to offer their programs at a larger scale.

Researchers

Researchers bring rigorous methodological training, global knowledge about their domain, experience from the history of their field, and creativity to map innovative programs to local problems. Researchers can also anticipate potential unintended consequences and general equilibrium effects of different programs and policies.

While most researchers would like to work on the most policy-relevant issues and see the work they were involved in scaled up, they can be constrained by the demands of academia, particularly the fact that top journals require cutting-edge research (Al-Ubaydli et al., Chapter 6 of this volume). Elsewhere, we discuss how to encourage research partnerships between governments and researchers (Dhaliwal and Tulloch, 2012). On the scaling side, however, for research to be scaled in the field, it must also be highly policy relevant and well aligned with the priorities of governments and implementers.

The constraints of researchers are also largely related to their teaching and research commitments. Scaling up programs may require extensive knowledge of government rules and regulations, established networks with local implementers, and significant time coordinating activities in the field. This is not the comparative advantage of many researchers, except perhaps for researchers who are interested in answering research questions specific to scale (see Mobarak & Davis, Chapter 12), and even then, they often need significant organizational support. Researchers may also benefit from a direct line of communication with on-the-ground representatives who can help them better understand local context.

Funders

Funders outside government, whether private foundations or multilaterals like development banks, bring many contributions to a collaboration in addition to providing funds to catalyze change. Funders often have a laser focus on outcomes, as grant requirements are often designed around specific metrics. This can provide accountability. They also often have sector expertise and experience that allows them to be knowledge partners. Finally, funders can support the growth and development of organizations who could benefit from capacity building and infrastructure improvements.

Incentives for funders often focus on creating the biggest impact for the money spent. This is also what motivates many of them to support research designed to understand not only the impact of programs but also the reasons underlying the change. This also means that scale-ups can be attractive to donors who want to ensure that their impact is wide reaching.

However, foundations may be constrained by rules that govern what types of programs they fund, or which sectors are priorities. While individual philanthropists are not bound by such constraints, they may often also have strong opinions on what works—or what doesn’t work—and thus may direct their funding in a less evidence-informed manner. As funders are often geographically distant from many of their target beneficiaries, they may need to rely on other collaborators for contextual knowledge.

E2P Organizations

Evidence-to-policy organizations (or E2P organizations, alternatively referred to as evidence translators,[ii] conveners, or knowledge partners) bridge the gap between research and evidence on the one hand and policy and scale-up on the other. J-PAL is one example of an evidence-to-policy organization. Our 400+ affiliated researchers at universities around the world design and implement field evaluations, many run by local J-PAL offices, while our policy and training staff build close partnerships with policy makers, NGOs, and funders. Other examples include (but are certainly not limited to) the World Bank, the Inter-American Development Bank (see Chapter 11 by Araujo et al.), and our partner organization Innovations for Poverty Action (IPA). By bringing together the other actors, and providing key inputs into research, policy, and capacity building, E2P organizations can reduce the constraints to collaboration, negotiate across the diverse perspectives and incentives of various partners, and ensure that evidence is translated into policies effectively. Although not always necessary, this role smooths the collaboration process and allows for stakeholders to take advantage of windows of opportunity. Given that E2P organizations often have less frequent transfers than governments, they can also serve as important sources of continuity and institutional memory when administrations or key officials change (Coburn et al., 2013).

To be effective, E2P organizations should have convening power, a globally-informed and locally grounded knowledge base, skilled full-time staff in policy outreach and communications, permanent local presence, deep experience in the field, an explicit mandate to bring together various collaborators to scale up evidence-based programs, and readily available resources to leverage policy windows. They are a non-partisan addition to a collaboration whose primary agenda is generating and increasing the use of evidence in policy design and decisions. E2P organizations' missions typically include encouraging the scale up of evidence-based programs. However, this relatively narrow mandate can be a constraint due to limitations that may exist on what types of questions they can answer or data they can use. Other constraints include limited funding or capacity relative to the scale of the problem and excess demand for their services. Finally, as these organizations often play the role of convening and catalyzing partnerships, they are inherently constrained by the aggregate constraints of partners with whom they work.

Forging Partnerships for Scale: How to Build Partnerships, Align Goals, Balance Competing Priorities, and Combine the Complementary Skills and Expertise of Each Actor

Understanding the incentives and constraints of various stakeholders and building partnerships across organizations are just as critical for a successful scale-up as the technical components of scaling, such as understanding the program's general equilibrium effects. Successfully scaling a program requires, among other elements, a deep understanding of global knowledge and local context and systems, a policy window where change is possible, political will to change the status quo, adequate funding resources, and capacity to monitor and implement the program well. Very often, all of this requires a coalition of organizations working together. In this section, we summarize four key lessons for forging strong scaling collaborations based on J-PAL's 15+ years of experience convening policy makers, NGOs, researchers, other E2P organizations, and funders to bring evidence-informed programs to scale, which to date have reached more than 400 million people worldwide (J-PAL, 2018).

Lesson 1: Long-Term Partnerships for Real-Time Responses

Many have emphasized the importance of long-term partnerships with governments in translating evidence into policy, and the same is true for scaling effective programs (Canales et al., 2018, Coburn et al. 2013, Cooley et al., 2016, Muralidharan et al., 2016, Penuel & Gallagher, 2017). All of the scale-ups to which J-PAL has contributed happened in the context of a multi-year partnership that included a mix of collaborative research, capacity building, and policy work.

For example, in 2013, the Indonesian government issued social protection ID cards to more than 15.5 million low-income households nationwide after a randomized evaluation conducted by J-PAL affiliated researchers and funded by the Australian government showed that the cards substantially improved the delivery of the country's largest social assistance program.[iii] At the time of the scale-up, this coalition of a government agency (Indonesia's National Team for Accelerating Poverty Reduction, TNP2K), researchers (Abhijit Banerjee, Rema Hanna, Jordan Kyle, Ben Olken, and Sudarno Sumarto), funders (Government of Australia), and E2P organizations (J-PAL) had already been collaborating on policy-relevant research for more than five years.[iv, v]

Similarly, the Indian NGO Pratham worked with J-PAL affiliated researchers, J-PAL's South Asia office, and schools across India for more than a decade to test and refine the pedagogical approach Teaching at the Right Level (TaRL), demonstrating that teaching to the level of the child, rather than a predetermined grade level, for a portion of instruction time can be enormously useful for helping children learn. Pratham has since scaled the program to reach more than 40 million children in India in partnership with funders and public and private schools in India. The approach is also being adapted to several countries in Sub-Saharan Africa. Together J-PAL and Pratham conducted five randomized evaluations testing Pratham's TaRL programs at different scales, implementation models, and with different implementers. This long-term collaboration helped identify key mechanisms that made the approach effective, the contexts where it could be most relevant, and identify multiple versions of the program that worked—whether delivered during or after school and by tutors, volunteers, government teachers, or through education technology—providing other governments and NGOs options for adapting the approach to their unique contexts and education systems (Banerjee et al., 2017).

Long-term partnerships are critical for scaling for many reasons. First, they foster trust and mutual understanding. Building a reputation among many people in a government allows us to identify policy makers who have the authority, ability, and interest to successfully scale a program (Andrews et al., 2010). And working with the same organization to solve multiple problems over several years demonstrates our commitment to building their capacity and helping them achieve their goals, not just our own.

Longer-term partnerships also make it easier to identify policy windows and respond to them quickly (Kingdon, 1995). While some windows are cyclical and predictable, like budgeting cycles and changes in administration following elections, many arise unexpectedly. Having a permanent presence on the ground and frequent in-person interactions with governments make it more likely that they will reach out to researchers or E2P organizations when these windows arise.

For example, we learned about the opportunity to evaluate Indonesia's social protection ID cards because one of the researchers was working with the government on another evaluation and learned that the Vice President requested evidence before committing to scale up the ID cards. Since we were already working together, we had the staff, agreements with government, and existing relationships with implementers and survey firms required to respond quickly and generate the evidence the Vice President requested in less than a year to inform their national budgeting process.

Since most scale-ups take several years, they also inevitably encounter changes in key personnel, entire administrations, and policy priorities. Longer-term partnerships give us time to build support among many people in government beyond the key champion and set up formal agreements that can institutionalize a scaling effort beyond an individual champion's efforts. Even if personnel don't change, government priorities will. Close, regular interaction allows each partner to understand how and why priorities have changed and to adapt the strategy and marketing of the scaling effort to new political realities.

Last, collaborating with governments or NGOs on multiple projects over many years allows for deeper learning and institutional change than a one-time project. Typically, the first project that researchers, NGOs, and governments work on is smaller and lower risk. If the first project is successful, governments are often more willing to test larger programs or consider scaling an effective pilot program they previously tested.

Challenges to Long-Term Thinking

While long-term collaborations are useful, they take time and resources to build and require each actor to operate differently than they normally would. The tenure of civil servants is sometimes only a few years. Researchers often only have incentives to work on a scaling effort if they can generate additional research projects with the same implementer. NGOs may not be able to provide governments with long-term support if their funders require concrete results within 2-3 years, so they may choose direct implementation instead of scaling through government infrastructure. Funders may not be able to support long-term scaling collaborations when the timelines are long and expected results are less certain.

Overcoming the Challenges. What can each collaborator do to forge longer-term partnerships in spite of their constraints? From all sides, it requires patience and a willingness to engage repeatedly despite our different incentives and to sacrifice some of our own short-term needs to achieve a larger goal. As Coburn et al. eloquently put it, this requires a "commitment to mutualism—sustained interaction that benefits both researchers and practitioners" (Coburn et al., 2013).

For governments, this means being open to innovation and learning throughout the design and monitoring of their programs, and having the patience to pilot, experiment, and obtain results before deciding to scale up.

For researchers, it means keeping scalability in mind from the beginning (Cooley et al., 2016) and being willing to test policy makers' priority interventions. It means generating evidence about program models that governments could actually scale, rather than perfect programs that can only be delivered on a small scale. It could also mean testing tweaks to existing large programs, which may be politically and administratively easier to scale than an entirely new program (Muralidharan et al., 2016; Duflo, 2017).

For E2P organizations, this requires having a permanent presence on the ground so that regular face-to-face interaction is possible. Local universities are thus well placed because they are deeply and permanently embedded in the context, science and evidence are core to their missions, and they have credibility with governments. However, many universities and civil society organizations in low- and middle-income countries lack the resources needed to provide this long-term support. Without flexible, long-term funding, local E2P organizations cannot devote time to building the valuable relationships that make them effective partners to the other collaborators involved in scale-up efforts.

International and local funders thus have an important role to play. Scaling efforts are typically financed by several different funders over their lifecycle. Private philanthropy or international development agencies often have more flexibility than a government to provide seed funding for the research and piloting that precedes a scale-up effort, whereas governments may be able to mobilize larger amounts of funding after a successful pilot.

Unfortunately, there is not enough philanthropic, aid, or government funding available for the earlier (riskier) stages of scale-ups. Many current scale-up funding sources take more than six months from application to funder response, while many promising policy windows do not last more than 2-3 months. Two types of funding are therefore particularly useful.

First, providing some flexible unrestricted support for effective E2P organizations helps cover the fixed costs of building longer-term partnerships where more meaningful reform and change at scale is possible. This allows them to have permanent staff who can regularly interact with policy makers and provide ongoing capacity building and mapping of policy priorities to global evidence.

Second, setting aside funding for high-quality scaling efforts, which provide decisions to applicant organizations within a few months, is critical for taking advantage of unexpected policy windows. For instance, we were only able to conduct the ID card evaluation in Indonesia on the government's timeline because the Government of Australia had a quick-response fund with a rolling deadline for policy-relevant impact evaluations with the Government of Indonesia.

At J-PAL, we accomplish this through our initiative structure, in which funds sit within specific regions or thematic areas to fund further research and policy partnerships. Initiatives begin with collaborative conversations between donors, policy partners, and researchers to (a) identify key policy priorities, (b) determine gaps in existing evidence on how to address those challenges, and (c) fund proposals for the design, pilot, evaluation, and scaling of innovative solutions.

In many countries, the government may be in a position to finance this type of fund. But more often than not, philanthropy and foundations are the biggest catalysts for this change.

Lesson 2: Incorporate Learning from Multiple Sources Throughout the Scale-Up Process

The path to scale is never automatic. Instead, the effort, enthusiasm, and active support of many collaborators are needed at every step along the way. In addition, many types of information from various sources are needed to make sure that the right programs are scaled up and that the process is as effective as possible. It is hard to scale down a program, so before decisions are made that will affect millions of people and cost many millions of dollars, it is crucial that policy makers consider several different types of data and evidence to inform the decision-making process.

In addition to what data is useful, we will also discuss when this data should be used (that is to say, continuously rather than just at the beginning or the end of a scale-up) and where it should come from (ideally, both global and local sources).

Types of Data

Rigorous Evidence of Effectiveness. Well-designed and implemented randomized evaluations are a particularly rigorous way of establishing causal links between a program and its results. Through randomization, the group assigned to receive a program is not systematically different from the group that was not. This allows us to compare the outcomes of the two groups in order to estimate the causal impact of a program. This is especially important to establish before a program is scaled.

When using evidence from randomized evaluations to consider supporting scale-ups, creating lessons that may be generalized to new contexts, or making other policy recommendations, we focus on the mechanisms of a program that led to its success—the *why*, rather than just the *what*. Strong partnerships and clear lines of communication are necessary to turn the data from a randomized evaluation into a reliable understanding of the mechanism behind the success of a program.

Descriptive Research on the Nature of the Problem and Scaling Opportunity. Descriptive research can be a useful addition to experimental research to prepare for the process of scaling up. Gathering simple statistics on the nature and extent of a problem in a context is critical for determining whether a program is relevant (Bates and Glennerster, 2017).

What are the key challenges or opportunities people are facing in this context? What do they want or need? Answers to these types of questions can often be gathered most easily through interviews, surveys, or focus groups.

Descriptive research is particularly vital during the early stages of collaboration. Before considering possible solutions, we first need to understand the nature and extent of the problem in a context, and whether an evidence-informed approach we have in mind is relevant for the problem at hand and implementable in that setting. For example, J-PAL and Pratham learned that the TaRL approach is more relevant and feasible to implement in contexts where children in grades 3-5 are not mastering foundational skills, learning levels are highly heterogeneous, and public schools have teacher support systems. In contexts where learning levels are high and most students are mastering basic skills, as we found in a descriptive research of a pilot of TaRL in Peru, the program may not be a relevant option.

Continuous Feedback from Participants and Frontline Implementers. Continuous feedback can and should be solicited throughout the entire process of program design and implementation. It is increasingly recognized that human-centered knowledge is an important aspect of research (See Lyon, Chapter 13).[vi] Continuously collecting participant feedback, Twersky (2019) explains, can demonstrate respect for participants' viewpoints and can provide contextually appropriate innovations that implementers may not generate themselves.

An example of the usefulness of direct participant feedback comes from the Indian state of Andhra Pradesh, where corruption and leakages resulted in money meant for anti-poverty programs not always reaching intended beneficiaries. One proposed solution was a biometrically authenticated "Smartcard" system for government-to-people payments. J-PAL affiliates Karthik Muralidharan, Paul Niehaus, and Sandip Sukhtankar conducted a randomized evaluation of a Smartcard system targeting pension and public employment payments covering more than 19 million people. The research showed the Smartcards were effective in both reducing corruption and improving program delivery: Government spending stayed the same, but beneficiaries reported receiving higher payments, suggesting that program losses had decreased by more than 40 percent.

Here's where beneficiary feedback was key to scale: 90 percent of beneficiaries reported preferring the Smartcards over the old system. Officials who had been in positions to gain from previous leakages, and who were now losing out, tried to push back by saying the program was inconveniencing people who now had to go through the process of matching biometric data to receive benefits (Muralidharan et al., 2016). But with feedback clearly demonstrating the public's preference for the new Smartcards, the government had a clear mandate to keep the program scaled up across the state.

Frontline program implementers can also provide useful feedback to help us learn more about on-the-ground experiences. What elements of a new curriculum are easy for teachers to implement, and which are not? Gathering this kind of direct input from the frontline service providers delivering the program (who, like participants, are important program users) is critical for scaling any program effectively—for ensuring it is not too burdensome to implement and that they can implement it consistently, and for generating their buy-in.

It is important to keep in mind how and when information will be used. Specifically, we must keep in mind that a program ready for scale is not set in stone. Continuous feedback necessitates a willingness and ability for continuous adaptation and tweaking. J-PAL's cofounder Esther Duflo refers to this concept as *plumbing* (Duflo, 2017). Plumbers, she notes, install machinery, observe, and then—critically—they tinker. This tinkering is often done in real time. Several adjustments may need to be made, without a guarantee that each will result in the best possible outcome. Tinkering is a process necessarily filled with a level of uncertainty. It can be messy.

Of course, tinkering with an idea, or incorporating user feedback, is not just a process to consider after a program has begun, but throughout the entire process. Often, this continuous engagement is part of a larger ongoing process of contextualization of a program to adapt to a location; as more is learned about the new context, implementation details must continually shift to fit.

Administrative Data. Finally, it is important to take advantage of administrative data. Every day, governments are collecting administrative data on all sorts of programs. The accuracy, quality, and availability of administrative data has greatly improved in the past decade and continues to improve as more is now digitally collected at the point of transaction, linked to biometric and other authentications, and transmitted to servers in real time. Existing systems already set up to collect information can be used as inputs at various steps. Primary data can be expensive and time-consuming to collect—remember to take advantage of data already being collected whenever possible. For example, in Philadelphia, J-PAL affiliated researchers are using administrative data to track the impacts of a summer jobs program on a large number of otherwise hard-to-follow outcomes—including crime, employment, and educational attainment.

Globally Informed, Locally Grounded. At J-PAL, we say programs should be globally informed and locally grounded. What we mean is that understanding local needs and formal and informal structures in a specific context is vital to the success of any program. But that doesn't mean global knowledge isn't useful: If knowledge from another context illuminates a general characteristic of human behavior—the mechanism behind why people behave in ways that make programs successful or particular technological solutions that are replicable in other contexts, such as machine learning algorithms—then we can use that knowledge to inform programs in other settings (Bates and Glennerster, 2017).

For example, to determine whether the Pratham's TaRL approach could also be relevant in Zambia, J-PAL's Innovation in Government Initiative (IGI) supported the journeys of officials from the Zambian Ministry of Education to travel to India to see Pratham implement TaRL firsthand, and supported Pratham staff to travel to Zambia to train the ministry in implementing the approach. The funding also enabled J-PAL Africa and Pratham to help the Zambian government adapt and pilot three versions of the program that were still consistent with the evidence, but allowed the government to learn which model best fit the needs and constraints of their school system. These journeys and piloting (along with plumbing, research, partnership building, and a variety of funding sources) eventually led to the successful scale-up of TaRL to 1,800 schools across Zambia.

These types of globally informed partnerships can only succeed, of course, with a local on-the-ground presence, the capacity of local institutions to carry out research and scale up new programs, and a favorable political context. Understanding the history of programs that have already been tried, as well as the priorities of the current administration, can assist various stakeholders in applying global evidence to a local context.

Lesson 3: Institutionalizing a Culture of Evidence-Informed Policy Making

In addition to—and often alongside—scaling up individual programs, there is often room to “scale up” or institutionalize a broader culture of evidence-informed policy making itself within governments, NGOs, and funders. Investing in capacity-building efforts that enable governments to adopt a broader culture of data and evidence use can lay the groundwork for more, and more effective, scale-up collaborations in the future. These efforts can also empower governments to fully own the scaling process. And institutionalizing an interest in evidence among funders can create more funding opportunities for scale-ups of evidence-based programs.

Three broad foundations lay the groundwork for making data and evidence use more feasible (Carter et al., 2018). First, policy makers within an institution must have the technical capacity to identify, access, evaluate the quality of, and apply data and evidence that is relevant to their policy goals. They must also be able to determine when existing evidence cannot address their particular questions, and to conduct or commission new research to fill such gaps (see Chapter 7, Ioannidis et al.).

Second, policy makers must have access to administrative data systems that enable continuous learning. Data must be stored, formatted, and analyzed in a way that makes it easy for policy makers to identify useful insights. Good data access and analysis systems can inform day-to-day program management decisions, significantly reduce the costs of conducting evaluations, and enable governments to monitor implementation. Unfortunately, many governments and NGOs do not yet have the technical capacity for effective data access and analysis. This is where E2P organizations have a key role to play. For example, J-PAL's Innovations in Data and Experiments for Action Initiative (IDEA) provides common guidelines, shared knowledge on best practices, and technical assistance to data providers and researchers to support safe and streamlined access protocols and high-quality data collection and documentation to enable new collaborations building on this data.

Finally, institutions must design systems and processes that encourage the use of data and evidence. While committed individuals can make evidence-informed decisions, creating an institutional culture requires formal systems, guidelines, and/or incentives that encourage the use of evidence in policy. These systems and processes can take many forms, ranging from prescriptive and resource-intensive on the one hand, to lighter-touch and more directional on the other. For example, the Department of Planning in the Indian state of Tamil Nadu created a dedicated fund for program evaluation of US$1.5 million per year in 2017. Alongside the creation of this fund, the Department also issued evaluation guidelines for impact evaluations of government programs (Evaluation and Applied Research Department, 2017). A less intensive example is the application criteria developed by the Chilean Ministry of Economy's Innovation Fund. Applicants to this competitive fund must include a theory of change and a review of existing evidence when seeking funds.

Supporting Champions to Build an Institutional Culture of Data and Evidence Use

How do governments develop these complementary foundations? In many cases, a motivated champion will push governments in the direction of greater data and evidence use. Champions within government are senior-level officials who have the autonomy to approve key decisions and who are invested in effectiveness. They are deeply committed to promoting a culture of evidence-informed policy making and often allocate time and resources outside of their regular commitments to furthering this goal.

Government champions are crucial for building this type of culture, but they are often transferred or pursue other opportunities following administration changes. Therefore, it is important that champions' personal efforts are not the only impetus for data use within institutions. Champions, often working closely with E2P organizations, can use a number of strategies to foster broader, institution-wide acknowledgement of the importance of data and evidence.

First, making evidence use (a part of) someone's job within the government can be a highly useful strategy for increasing data and evidence use. In many cases, funding from external sources can enable the appointment of embedded staff from E2P organizations to sit within the government and focus on evidence use. Embedded staff make important contributions to long-term collaborations crucial for effective scaling collaborations. They build trust and mutual understanding between the government and its partners, and they are well-placed to identify policy windows.

Where it is possible to go beyond reserving staff time for evidence use, governments can build dedicated internal institutions that focus on innovation, evaluation, and the application of evidence. Similar to the "plan, do, study, act" cycles of improvement science or continuous improvement (Bryk et al., 2015), these departments often institutionalize an iterative "learning cycle" (Carter et al., 2018) of diagnosing the problem, designing and piloting an intervention based on existing evidence, evaluating it, and then feeding the results directly into policy, including scaling the program if it is found effective. The process then begins again with a different issue. These institutions create incentives and safe spaces for policy makers to propose new ideas and help build an understanding of data and evidence use as tools for learning and improvement, rather than for accountability. In Peru, for example, the Ministry of Education worked with J-PAL and IPA to establish MineduLAB, a unit to identify, test, and scale low-cost interventions to improve educational outcomes. MineduLAB scaled up its first innovation in 2017, reaching over 500,000 children, and is currently scaling two additional innovations that were evaluated using administrative data and found to be effective. MineduLAB has survived multiple administration changes in Peru and has inspired the creation of several other innovation and evaluation hubs.[vii]

Governments can work with collaborators to build the foundations that enable an institutional culture of data and evidence use. Throughout scale-up collaborations, researchers and evidence-to-policy organizations should create opportunities for government staff to engage with, lead, and co-generate activities involved in scaling up a program so they are equipped to engage in these activities once the collaboration has ended.

Lesson 4: The Critical Role of E2P Organizations

Throughout this chapter, we have referred to evidence-to-policy organizations (E2P organizations), or the collaborators whose role is to ensure that existing evidence is applied to policy decisions. Since their mission is to make sure that policy is informed by evidence, they have a mandate to encourage the scale-up of evidence-informed programs.

Evidence-informed scale-ups can and do happen without an E2P organization. But bringing each of these collaborators together to work toward one goal can be difficult if relationships are not already established. The strengths, incentives, and constraints of each partner are not necessarily immediately compatible, meaning that a partnership might not be an obvious fit at first. E2P organizations can be key catalysts to overcome these obstacles and help build systems that apply evidence to policy making and then fund and support implementation.

In order for E2P organizations to manage these often tricky relationships, certain factors must hold. Core funding is necessary. Employing staff in the field is important to ensure that the organization maintains local relationships and has the capacity to understand local contexts. There must also be demand: E2P organizations only have a role to play if there are collaborations in need of support in generating or using evidence, securing funding, government backing, or implementation and monitoring expertise.

In return, E2P organizations allow for flexibility to respond to new constraints and challenges as they arise. If a new problem demands an innovative response, E2P organizations can review and synthesize evidence and convene researchers to provide solutions. If funding from one source dries up, E2P organizations can call on other funders whose interests align with the project. If one implementation channel is no longer viable, E2P organizations may have relationships with others that have worked well previously. E2P organizations can review evidence gaps and provide relevant evidence, either summarized from a relevant study or synthesized across many. E2P organizations conduct cost-effectiveness analyses to ensure money is being spent wisely (Dhaliwal et al., 2011). E2P organizations can also help set up monitoring and evaluation systems and use administrative data for understanding context and then for monitoring feedback loops from an independent and unbiased perspective.

E2P organizations may also improve the chances that a program can be scaled up while maintaining fidelity to the most important features critical in the success of the original tested model, while being able to advise on which elements can be modified, adapted, or dropped without affecting program impact. Without an organization whose incentive is to ensure that evidence remains front and center, scale-up processes may stray far from what was tested in order to be more convenient or feasible for one partner or another.

E2P organizations' experience with scale-up collaborations also means that they are ideally placed to provide capacity building and training for many of the other players in a partnership. While researchers might need help training research staff, funders might appreciate knowledge-building to better target their support, while implementers or government partners may benefit from capacity-building initiatives that allow their staff to better interpret and use evidence. E2P organizations often specialize in capacity building to generate and use evidence. For example, J-PAL provides in-depth online and in-person trainings on evidence generation and use, designed to meet each partner at their current level of need and directly support their growth.

Ultimately, the importance of E2P organizations lies in their singular focus on applying evidence to policy at scale. Although collaborations can of course happen without a convener, E2P organizations can play a vital role in innovation, generating rigorous evidence, relationship-building, and institutionalizing systems for evidence to scale.

Conclusion

Scaling evidence-informed programs is equally science and art. While it is important to understand the theoretical and economic underpinnings of scaling, the true test is when scale-up projects meet the complexities and realities of the real world. When governments, practitioners, researchers, and funders build long-term partnerships, often with the support of an E2P organization, each becomes part of a larger ecosystem designed to improve the lives of many more people than could be reached by any of these actors alone. Their complementary skills, and the interaction among each actor's unique incentives and constraints, allow the whole to be greater than the sum of its parts.

References

Abdul Latif Jameel Poverty Action Lab (J-PAL). (2018, November 8). Evidence to Policy. Retrieved March 15, 2020, from https://www.povertyactionlab.org/evidence-to-policy

Al-Ubaydli, O., List, J. A., and Suskind, D. L. (2019). The science of using science: Towards an understanding of the threats to scaling experiments. (NBER Working Paper No. 25848). Retrieved from the National Bureau of Economic Research website: https://www.nber.org/papers/w25848

Andrews, M., McConnell, J., and Wescott, A. O. (2010). Development as leadership-led change – A report for the Global Leadership Initiative and the World Bank Institute (WBI). (HKS Faculty Research Working Paper Series No. RWP10-009). Retrieved from Social Science Research Network website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1551375

Banerjee, A., Hanna, R., Kyle, J., Olken, B. A., and Sumarto, S. (2017). The role of competition in effective outsourcing: Subsidized food distribution in Indonesia. (Working Paper). Retrieved from J-PAL website: https://www.povertyactionlab.org/publication/role-competition-effective-outsourcing-subsidized-food-distribution-indonesia

Banerjee, A., Hanna, R., Kyle, J., Olken, B. A., and Sumarto, S. (2018). Tangible information and citizen empowerment: Identification cards and food subsidy programs in Indonesia. *Journal of Political Economy, 126*(2), 451-491. https://doi.org/10.1086/696226

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics, 122*(3), 1235–1264. https://doi.org/10.1162/qjec.122.3.1235

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R., and Khemani, S. (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy 2*(1), 1–30. https://doi.org/10.1257/pol.2.1.1

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M., and Walton, M. (2016). Mainstreaming an effective intervention: Evidence from randomized evaluations of 'Teaching at the Right Level' in India. (NBER Working Paper No. 22746). Retrieved from the National Bureau of Economic Research website: https://www.nber.org/papers/w22746

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., and Walton, M. (2017). "From proof of concept to scalable policies: Challenges and solutions, with an application." Journal of Economic Perspectives 31, no. 4: 73-102. https://doi.org/10.1257/jep.31.4.73

Bates, M. and Glennerster, R. (2017). "The generalizability puzzle." *Stanford Social Innovation Review 15*(3), 50-54.

Bryk, A., Gomez, L., Grunow, A., and LeMahieu, P.. *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press, 2015. https://www.carnegiefoundation.org/resources/publications/learning-to-improve/

Canales, R., Sheldon, T., Grieves, T., and Lehan, B. (2018). Evidence in practice: Toward an integrated approach. Yale School of Management. https://evidenceinpractice.yale.edu/full-report/

Carter, S., Dhaliwal, I., Katticaran, J., Macías, C., and Walsh, C. (2018). Creating a Culture of Evidence Use: Lessons from J-PAL's Government Partnerships in Latin America. https://www.povertyactionlab.org/gpi/creating-culture-evidence-use

Coburn, C., Penuel, W., and Geil, K. (2013). "Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts." *New York, NY: William T. Grant Foundation*. https://eric.ed.gov/?id=ED568396

Cooley, L., Kohl, R,and Ved, R. R. (2016). *Scaling up—from vision to large-scale change: A management framework for practitioners*. Third Edition. Management Systems International.

Dhaliwal, I., and Tulloch, C. (2012). From research to policy: using evidence from impact evaluations to inform development policy. *Journal of Development Effectiveness 4*(4), 515-536. https://doi.org/10.1080/19439342.2012.716857

Dhaliwal, I., Duflo, E., Glennerster, R., and Tulloch, C. (2014). Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education. In P. Glewwe (Ed.), *Education Policy in Developing Countries* (pp. 285-338). The University of Chicago Press.

Duflo, E. (2017). Richard T. Ely Lecture: The Economist as Plumber. *American Economic Review 107*(5), 1-26. https://doi.org/10.1257/aer.p20171153

Evaluation and Applied Research Department. (2017). *Evaluation and Applied Research Department – Institutionalize an Evidence-Based Approach to Policy Making – Approval of the Guidelines for Programme Evaluations – Orders – Issues* (G.O.(Ms).N0. 80). Government of Tamil Nadu. https://www.direar.tn.gov.in/adm/writereaddata/upload/evaluation_related/Evaluation%20relelated%20GOs.pdf

Kingdon, J. W. (1995). *Agendas, Alternatives, and Public Policies*. HarperCollins College.

Muralidharan, K. and Niehaus, P. Experimentation at scale. *Journal of Economic Perspectives 31*(4), 103-24. https://doi.org/10.1257/jep.31.4.103

Muralidharan, K., Niehaus, P., and Sukhtankar, S. Building state capacity: Evidence from biometric smartcards in India. *American Economic Review 106*(10), 2895-2929. https://doi.org/10.1257/aer.20141346

Penuel, W., and Gallagher, D. *Creating Research Practice Partnerships in Education*. Harvard Education Press. 8 Story Street First Floor, Cambridge, MA 02138, 2017.

https://www.hepg.org/hep-home/books/creating-research-practice-partnerships-in-educati

Poirrier, C., Heller, N., and Whitt, P. 2018. "Scoping Study: Evidence Translators' Role in Evidence-Informed Policymaking." *Results for Development.* https://www.r4d.org/resources/scoping-study-evidence-translators-role-in-evidence-informed-policymaking/

Twersky, F. (2019). Time for a three-legged measurement stool. *Stanford Social Innovation Review*.

Notes

[i] See J-PAL's Evidence to Policy page (https://www.povertyactionlab.org/evidence-to-policy) for a library of case studies of collaborations to scale evidence-informed programs around the world.

[ii] Poirrier, Caroline, Nathaniel Heller, and Preston Whitt. 2018. "Scoping Study: Evidence Translators' Role in Evidence-Informed Policymaking." *Results for Development.* https://www.r4d.org/resources/scoping-study-evidence-translators-role-in-evidence-informed-policymaking/

[iii] J-PAL. "Case Study: Targeted information to improve social assistance." Updated 2017. https://www.povertyactionlab.org/case-study/using-identification-cards-improve-national-social-assistance-indonesia

[iv] Banerjee, Abhijit, Rema Hanna, Jordan Kyle, Benjamin A. Olken, and Sudarno Sumarto, 2018. "Tangible Information and Citizen Empowerment: Identification Cards and Food Subsidy Programs in Indonesia." Journal of Political Economy 126 (2): 451-491. https://doi.org/10.1086/696226.

[v] Banerjee, Abhijit, Rema Hanna, Jordan Kyle, Benjamin A. Olken, and Sudarno Sumarto. "The Role of Competition in Effective Outsourcing: Subsidized Food Distribution in Indonesia." Working Paper, March 2017. https://www.povertyactionlab.org/publication/role-competition-effective-outsourcing-subsidized-food-distribution-indonesia

[vi] Twersky, Fay. 2019. "Time for a Three-Legged Measurement Stool." *Stanford Social Innovation Review* Winter 2019.

[vii] For more information on MineduLAB, see J-PAL's Evidence to Policy case study, "A government innovation lab to improve education."