Evidence in practice: how structural and programmatic scaffolds enable knowledge transfer in international development

Evidence in Practice: How Structural and Programmatic Scaffolds Enable Knowledge Transfer in International Development

Rodrigo Canales* Boston University
Mikaela Bradbury Goldman Sachs
Anthony Sheldon Yale University

February 8, 2023

PLEASE DO NOT CITE OR DISTRIBUTE

Abstract: This inductive study of eight international development interventions across four countries analyzes the mechanisms that enable effective integration of evidence in practice, as an enduring challenge of learning and coordination across occupational and organizational boundaries. We identify how a set of critical, complementary structural and programmatic scaffolding practices, jointly provided a shared architecture for actors across organizations and communities of practice to collaborate and learn within the uncertainty and complexity inherent in the international development context. Scaffolding practices offered modular, temporary counter-balancing stabilizing and extending support that enabled the unusual and counternormative behaviors and mindsets required for continuous learning and adaptive coordination. Through 226 in-depth interviews with a range of international development experts, including practitioners engaged in eight matched interventions in India, Mexico, South Africa, and Ghana, we identified and analyzed the mechanisms that explain the varying degrees of effectiveness with which rigorous evidence was integrated in each case. Our findings have implications for interorganizational innovation and collaboration under conditions of complexity and uncertainty, as well as the dynamic interactions between individuals, their organizations, and their communities of practice when attempting to bring about systemic change.

Introduction

There has been increasing discussion in practitioner, policy, and academic circles about the importance of "evidence-based practice" for international development work. The broader goal of evidence-based practice is to enhance efficacy, accountability, and learning in the growing field of development as it has in other fields, such as medicine. Ideally, decisions should be guided by empirical, rigorous, and objective knowledge rather than political, ideological, or other subjective factors. This is consequential because of the scale and urgency of the problems that development work seeks to address, as well as the amount of resources devoted to it. In 2019, USAID alone devoted $39Bn to development programs, while the OECD DAC spent $153Bn on international aid. And programs impact millions of people. IDA18 sought

*Corresponding author: rcanales@bu.edu

to reach 400 million people through nutrition programs, immunize 180 million children, train 10 million teachers who teach 300 million children, and bring safer childbirth to 20 million women. Across public and private domains, every dollar spent on a program that is not backed by nor generating useful evidence is a dollar potentially not spent on what actually works. The cost is borne twice: investments are made in interventions that are not necessarily the most effective and less evidence is generated to inform future decisions (Banerjee et al. 2011).

The nature and scale of challenges in international development imply that interventions occur at the intersection of different organizations and sectors. Programs often require coordination between policymakers from national and local governments, implementers from international and local NGOs, program officers from funding organizations, and researchers from think tanks and academic institutions. In addition, most development issues operate at the system level, where causality is difficult to determine, many factors are simultaneously at play, and cycles between action and impact can be non-linear, indirect, and prolonged. The use of public and donor funds in activities that affect vulnerable populations also creates important accountability concerns and a necessary focus on effective task execution, which is often at odds with generating and sharing the types of evidence that would allow for true learning (Gertler et al. 2016; Hausmann 2016). These factors hinder not only the generation of evidence, but also the clear identification of what should be considered as evidence for which parts of the system.

Evidence-based practice is thus a particularly clear and important instantiation of learning across communities of practice (Carlile 2004), which requires coordination and management of knowledge across organizational boundaries. Research has identified a variety of mechanisms that can help overcome a series of fundamental challenges to learning across communities of practice. These mechanisms—and associated challenges—include, for example, tangible boundary objects that allow for the identification and clarification of tacit misunderstandings that would otherwise derail a collaboration (Bechky 2003b,a; Carlile 2004); an established *project trajectory* that can help align and coordinate the disparate ways in which each community of practice would expect to sequence activities and outputs (Faraj and Xiao 2006; Fayard et al. 2017); relational structures that allow actors to build trusting and balanced relationships across boundaries despite the natural mistrust generated by conflicting values or priorities (DiBenigno 2018); team scaffolds (Valentine and Edmondson 2015) and trading zones (Kellogg et al. 2006) that create free spaces for negotiation between occupational groups with otherwise misaligned incentives and interests; and project spaces that allow actors to temporarily suspend and reframe their occupational identities to allow for different definitions of knowledge (Lifshitz-Assaf 2018). Finally, as problem or sys-

tem complexity increases, hierarchy can play a critical role in providing coherence of learning and clarity of vision amidst the diverging ethos of different communities of practice (Valentine 2018).

Collectively, these studies provide a helpful blueprint for the types of behaviors that help actors across communities of practice identify and overcome barriers to learning. Notice, however, that research has focused on collaborations across communities of practice that still reside within the same organization. In each of the documented cases, participants clearly had to stretch beyond their habitual mindsets, habits, and practices, but they were already collectively invested in a joint vision, an overarching goal, or the creation of a joint "product" (even if they might disagree on how to achieve that). Put differently, the organization—and how it has addressed the fundamental questions of organizing, such as how to define tasks and coordinate effort (Puranam 2018)—has been taken for granted. It is telling, for example, that relatively basic existential questions for a collaboration, such as sources and sustainability of funding, employment stability, authority over final decisions, or distribution of value do not seem to play a prominent role in the documented efforts. Empirically, the focus makes sense, as it has brought into full relief the individual (differences in) behaviors and departures from habitual patterns that allow actors to collaborate *conditional on embarking on a joint project.* But it has also left in the background the organizational and structural preconditions that made those—clearly challenging and risky—behaviors possible or desirable in the first place.

In many complex collaborations, such as development settings, a shared organization cannot be taken for granted. Actors must coordinate knowledge not only across communities of practice, but also across a diverse system of organizations and stakeholders. Each occupation and each organization may have a radically different understanding of what is at stake, not only in what type of learning is desirable, but also in the basic definition of goals and how to pursue them. Absent the backdrop of a common organization to provide a joint vision, basic stability, and clarity of authority (Valentine 2018; Zuckerman 2010; Puranam 2018), actors must stitch together basic structural preconditions and expectations *across organizations* and communities of practice to allow for coherence, stability, and synchronized learning. A better understanding of how those structural conditions and expectations are established has clear implications for cross-organizational collaborations like evidence-based practice. But it can also reveal the conditions, agreements, and expectations required to promote interdisciplinary learning that within-organization studies cannot easily observe.

In this paper, we study what people on the ground do to achieve learning across organizational and occupa-

tional boundaries in the complex development system. Through an inductive study of eight development interventions across four countries, complemented by 65 interviews with development practitioners and experts (for a total of 226 interviews), we analyze the practices that allow actors to sustain the collective and distributed learning necessary for integration of evidence in practice. We explore the mechanisms that enabled evidence integration and, more importantly, the underlying structural and programmatic conditions that allowed certain actors, and not others, to activate these necessary mechanisms. We identify a set of scaffolding practices that provide temporary support, guidance, and stability for actors. They are scaffolds because, while constituting a clearly structural element and approach, they are also temporary, modular, and interdependent with other elements at different stages of project design and execution. In particular, we identify a set of key *structuring* and *programmatic* scaffolding practices, which simultaneously extended and stabilized professional and personal norms around questions of identity, purpose, accountability, and knowledge. The identified scaffolding practices provided a set of guardrails that supported and guided actors towards knowledge transfer across their organizations and communities of practice, so they could navigate the uncertainty and complexity inherent in the international development context.

Through an integrated model of the scaffolding practices observed and the dynamics between them, our paper makes several contributions to theory and practice. First, while existing research has identified several practices, each tends to be described in isolation as an antidote to a particular challenge. Our research, in contrast, shows that scaffolding practices need to be deployed in concert, as it is their dynamic integration that provided the necessary support to sustain collaborations through their varied stages. What emerges, then, is a process theory of collaboration across boundaries, where complementary scaffolding practices provide the different types of support required at different stages of learning and development. Throughout this process, we highlight the importance of adaptive financing, a structural scaffolding practice that provides necessary financial security balanced with flexibility, adaptability, and a focus on learning. This is important, as financing concerns, while evidently central in all the cases we studied, play little to no role in prior research. Ultimately, many of the specific behaviors that allowed for successful learning across boundaries in our cases are quite consistent with prior research. We show, however, that these behaviors require actors to extend significantly from their existing routines, habits, and norms. In consequence, to sustain such extensions actors require the stabilizing functions provided by the integrated scaffolding practices we document.

Knowledge Coordination in International Development

Discussions of evidence-based practice frequently center on questions of codification and translation of knowledge (usually from researchers to practitioners) or the identification and replication of interventions that have been demonstrated, through rigorous research, to have an actual impact (Banerjee et al. 2011). This view tends to focus on the characteristics of the knowledge itself—the types of translation and codification that need to happen for it to become abstract and transferable enough to travel across contexts (Carlile 2004; Brown and Duguid 1991).

Research on the management and coordination of knowledge across boundaries, however, has long demonstrated the situated, social, participative, and negotiated nature of knowledge. In particular, knowledge is situated within communities of practice, which construct occupational mandates and jurisdiction over specific problems and appropriate ways of addressing them (Brown and Duguid 2001; Carlile 2004; Orlikowski 2002). Each community develops a distinctive ethos and identity, reflected in their values, practices, norms, symbols of meaning, and language (Lifshitz-Assaf 2018; Fayard et al. 2017; Bechky 2003b; Brown and Duguid 2001). This creates significant interpretative and structural barriers to coordination and management of knowledge across communities of practice (Carlile 2004). As the complexity of a system increases, the challenges of cross-boundary learning become amplified by inter-dependencies and the uncertainty of all that is not known about a complex problem, the system it resides in, and their shifting nature in time (Valentine 2018; Brown and Duguid 1991; Faraj and Xiao 2006).

Research has also shown that, to overcome these challenges, actors at the boundaries must redefine, negotiate, and transform knowledge to create a common ground (Carlile 2004; Bechky 2003a). This requires challenging, counternormative, and risky behaviors to reconfigure how actors perceive themselves (Lifshitz-Assaf 2018), how they relate to others (Kellogg et al. 2006; DiBenigno 2018; O'Mahony and Bechky 2008), and how they understand their (collective) purpose (Faraj and Xiao 2006; Dougherty 1992; Obstfeld 2012; Valentine and Edmondson 2015).

To facilitate such challenging reconfigurations, actors can deploy organizing structures that are structured enough to provide clear responsibilities and mechanisms for accountability, yet also remain flexible enough to allow for informal coordination, improvisation, and course corrections (Brown and Duguid 1998; Faraj and Xiao 2006). For example, a taskforce or project team can help create temporary, flexible, and clear guardrails that integrate actors across boundaries, delineate a collective focus, and delimit potential risks

(Obstfeld 2012; Kellogg et al. 2006; Faraj and Xiao 2006). They are only effective, however, when actors synchronize their interdependencies, identify and negotiate mutual obligations, and create mechanisms of accountability (Brown and Duguid 1998; DiBenigno 2018; Valentine and Edmondson 2015; Valentine 2018). This requires not only clarity in the collective vision and goals, but also a shared, abstract view of the entire system and how each actor can affect and be affected by others (Valentine 2018; Weick and Roberts 1993).

Research to date, in summary, has presented clear and useful guidelines for the specific behaviors that actors engage in to coordinate knowledge across communities of practice. It has also demonstrated that the process is difficult, contentious, risky—and accordingly relatively rare. And since the focus has been on the challenging and unusual behaviors that allow actors to collaborate across communities of practice, the structural preconditions that make those behaviors desirable, likely, or effective mostly remain in the background. Collaborations across communities of practice—and the behaviors that can render them successful—are challenging precisely because they depart from some of the norms, routines, and practices that help bring order to organizational and professional life. But the broader organization and the preconditions that made a collaboration desirable and possible in the first place remain. Actors still have clarity on what they are departing from—and will go back to—for what purpose, and with what organizational mandate.

More broadly, organizations are the result of a series of contracts or agreements that provide shared clarity on a series of fundamental questions that pervade all collective action (Coase, Gibbons, etc.). These start from the basic questions of what we want to accomplish together, how we define and distribute tasks, and how we coordinate effort. They help specify a collective goal and purpose, sources and uses of financing, the distribution of decision and design rights, the nature and stability of employment, mechanisms to settle disputes, and the distribution of value, information, and rewards (Puranam 2018). As actors under a shared organization seek to collaborate across communities of practice, they must extend beyond what they normally do, but they can still rely on the shared agreements and expectations provided by the organization as a backdrop that delimits uncertainty and allows them to focus on the challenging behaviors at hand. In other words, they can take the organization, and all that it provides, for granted.

For example, it is in the nature of learning across communities of practice that misunderstandings and disagreements will emerge. For more complex problems that can render such disagreements irreconcilable, hierarchy can become instrumental to provide clarity of purpose, a coherent understanding of

collective learning goals, and a clearing house for disagreements that would otherwise derail a collaboration (Valentine 2018). But hierarchy—a particular distribution of voice, decision, and design rights—is one of several features or agreements of a given organizational design, and it hinges on a variety of complementary features (Puranam 2018; Turco 2016).

By virtue of the diversity of communities of practice that must collaborate, we can anticipate that actors in a setting like international development will encounter the challenges described above and will need to activate the prescribed behaviors to overcome them. Because they also straddle sectors and organizations, however, there is no shared organization to fall back on. Even if actors may converge around a broadly defined issue or a given population of intended beneficiaries, each remains beholden and accountable to both her community of practice and her home organization. Absent a shared organization and the tacit and explicit agreements that it has codified, what is the baseline set of agreement that actors would need to stitch together to enable the challenging behaviors required for cross-boundary collaboration? And what process would they need to follow?

The lack of a unifying organizational identity creates a host of related challenges. First, actors across boundaries must establish a common ground and purpose, which will require them to relax attachments to the *how* of work—as defined by their organizations and communities of practice (Lifshitz-Assaf 2018). Second, and related, even if actors across boundaries are able to construct a common ground, each will remain attached to the expectations, performance evaluations, incentive systems, and measures of worth of their respective organizations and communities of practice (Carlile 2004; Kellogg et al. 2006). Yet, a fruitful collaboration requires more flexible and inclusive measures of value that can integrate the diverse set of perspectives, roles, and approaches that are at play, while still unifying them towards a shared outcome. How can actors relax their attachment to their work as defined by their habits and home organizations to justify and sustain the (substantial) boundary and knowledge work required to collaborate with others?

Third, absent a shared organization, actors will require new, temporary agreements that address some of the basic questions of organizing before they can deploy the behaviors required for collaboration. Given the lack of a formal structure, agreements will necessarily run on trust. A critical question, therefore, is how to create, maintain, and institutionalize trust across *organizations*—particularly given that trust is built through relationships between *individuals* (Sorenson and Rogan 2014; Canales and Greenberg 2015). Taken together, these challenges translate into the need for processes that provide accountability

and clarity of direction while also allowing for trusting relationships, flexibility, and dynamic adaptation of goals (Lingo and O'Mahony 2010; Davis and Eisenhardt 2011; Davis 2016).

Data and Methods

The research design consisted of three broad components. In the first, we conducted a set of 31 "expert" interviews with individuals who had spent a significant portion of their professional lives attempting, researching, or promoting the integration of evidence into development practice.[1] This included academics, government officials, foundation program officers, NGO practitioners, and think-tank directors. Table 1 shows the distribution of our interviews. To identify experts, we first contacted individuals who had either published extensively and prominently on the topic or who had actively funded research or programs with the explicit goal of integrating evidence into practice. From this first set of experts we conducted snowball sampling until we reached a saturation point.[2] This initial set of interviews informed and directed the next two components, as they resulted in an initial map of the relevant stakeholders in the "evidence-to-practice ecosystem" and the hypothesized and actual paths that seemed to link them together.

Second, we conducted a matched comparison of eight cases of development programs or interventions where rigorous evidence was integrated with varying degrees of effectiveness, as can be seen in Table 2. These cases were matched on structural, geographic, and programmatic characteristics –as well as on the extent to which evidence had informed practices– to better identify the critical factors that allowed actors in certain cases, and not others, to integrate rigorous evidence into practice. This matching process led us to identify pairs of cases across four different countries, leveraging temporal and cross-sectional variation between them as follows:

For each case, we first identified, through existing literature and interviews with subject experts, a series of key informants who had detailed knowledge of the case's history and protagonists. These initial interviews with case experts led to the creation of a detailed actor-stakeholder map for each case, where we identified the key stakeholder groups that either participated in or were affected by the program, as well as the specific individuals who played an active role in the program's evolution (self-identifying citation removed). These stakeholder maps were validated with several informants for each of the cases. We then conducted interviews with each of the key individuals, with a special focus on ensuring informed and representative views from every stakeholder group. Interviewees were asked to relate chronologies

of objective events, behaviors, choices at critical junctures, and facts of the processes described (Davis and Eisenhardt 2011). In every instance, the goal was to identify the individuals responsible for the particular evolution of a case, as well as the specific tactics they employed throughout the process, to better understand the rationale behind their decisions as well as the factors that led them to succeed or fail. In total, we conducted 161 interviews across the eight cases. Interviews were complemented with a wealth of archival information including media articles, private documents (donor reports, internal presentations and communications, etc.), and public documents (announcements, academic articles, editorial pieces). These data were used to trace the chronological list of events for the overall development of each case. Each storyline was developed in an extensive document that established the causal links described by the subjects and ensuring a balanced consideration of different stakeholders (George et al. 2005, Ch.5).

The third component consisted of interviews, outside of the scope of the case studies, with prototypical representatives of each of the stakeholder groups, or individuals who could clearly describe the *typical* experience of enacting a particular stakeholder role, as described in Table 1. Using the stakeholder map and initial hypotheses as starting points, this stage focused on the dynamics that shape the interactions between stakeholder categories. The work consisted of 34 in-depth interviews with representative actors from each stakeholder group, identified through our first set of interviews. The interviews focused on each individual's needs, assumptions, operational constraints, main concerns, professional and ideological backgrounds, timelines, and aspirations –especially concerning the development, dissemination, and use of novel evidence in development practice. This in-depth analysis resulted in a more nuanced stakeholder and system map that more clearly identified breakdown points and paths of connection that hinder or facilitate the exchange of knowledge across stakeholder groups. It also resulted in a refined set of hypotheses about the breakdown of communication and possible interventions to address it.

Across the three components, we conducted a total of 226 interviews. All interviews were in-depth and semi-structured, with an average length of around 90 minutes. Around two-thirds of them were done in person and the rest were conducted remotely. All interviews were recorded and transcribed verbatim.

Data Analysis

Data analysis was conducted in several stages. Each of the 226 interview transcripts was coded extensively to identify first-order concepts related to the integration of evidence into development practice. First-order concepts include "concerns about reputation" or "short-term decision-making." This required multiple readings of interview transcripts, field notes, and archival data to associate nearly every passage

of text with one or more codes. These codes were then grouped into second-order themes (Glaser and Strauss 1999), always contrasting them with current research on knowledge transfer. Second order themes included "incentive structures" or "timing misalignments," each of which was developed extensively in a memo that explored the characteristics, tensions, and contradictions of each theme. In stage three, we mapped the codes to each of our case narratives to detect patterns of activities, constraints, and decisions that defined the evolution of each case at critical junctures. This allowed us to identify similarities and discrepancies across cases, as well as to create comparable counterfactuals that could account for differing outcomes.[3]

In stage four, we created process maps, concept maps, data tables, and detailed case synopses that linked key challenges, events, and decisions to the specific alternative tactics employed by actors and then to their subsequent consequences for the development program or intervention in question. This final set of analyses revealed a somewhat consistent set of factors faced at comparable stages by actors across our different settings. Throughout our analysis, we iterated between emerging insights, existing theory, and matched comparisons across cases to identify the mechanisms that operated at critical junctures.

It is worth mentioning that, at two junctures of the project (the first after our first set of expert interviews was over and the second after the completion of our initial case narratives) we hosted workshops with two different groups of highly experienced representatives from each of the stakeholder groups. During these workshops, we discussed our emerging findings and we gathered additional, essential insights from participants. The workshops served to validate and deepen our understanding of emerging insights.

The Markers of Effective Integration of Evidence –Categorizing Success

We assessed the extent to which an intervention had integrated evidence in the same way that our informants did, relative to how existing evidence was identified and used, and the quality, quantity, and relevance of new evidence generated. In general, actors first sought to identify and synthesize (a diversity of) existing sources of evidence into an understanding of the development problem and context. Second, actors brought that understanding into the process they followed to design and implement the stages of an intervention. Critically, during implementation, *new* evidence emerged, that could both inform a program's own development as well as other, future interventions. The integration of evidence thus meant to identify what had worked, to do more of it, and to do it in more places. It also meant that participants became systematically better at the process of identification, adaptation, and transfer of evidence. Following this logic, integration of evidence could take three, not mutually exclusive, modes or

paths of effectiveness:

- *Scale:* The extent to which an intervention was improved and expanded in its intended settings both by learning from its own processes and by bringing in relevant knowledge from other programs.
- *Spread:* The extent to which the intervention generated new evidence which was then adapted and applied to new organizations and contexts.
- *Structure:* The extent to which, by virtue of their participation in the intervention, actors increased their organizational capacity to achieve scale and spread over time—or the extent to which they enhanced their organizational capacity to systematically generate, analyze, and integrate evidence.

As indicated in Table 3, there were clear differences in the effectiveness of evidence generation and integration across the eight interventions studied. We scored each intervention, using a 0-2 scale (not at all, partially, significantly), on the extent to which they had achieved the three paths of integration—scale (Sc), scope (Sp), and structure (St). From this categorization, our analyses contrasted more vs. less successful cases to better understand what could account for the differing outcomes.

Case Presentation

Our analysis drew extensively and equally from all our cases and interviews. However, given the complexity of the cases, in the text of the paper we will draw examples and quotations from only two of our eight cases to provide a clearer narrative thread. The tables and models still contain the data from all our cases for comparison, and are a necessary complement to all our written arguments. The pair of contrasting cases from India show one intervention intentionally designed and implemented to generate and integrate evidence into its ongoing practices (Pratham's signature program, Teaching at the Right Level) and a second intervention in which evidence generation and application were severely limited due to a misplaced conviction that the key uncertainties had been identified and addressed in the program's initial design (TARA's Aqua+). We selected this pair of cases because, as a matched pair, they both occurred within the same context and from broadly similar starting points. Furthermore, as a pair, they provided the clearest contrast in their chosen paths, across almost all dimensions of the process model that emerged from our analyses, thus allowing us to highlight practically every aspect of our process model without resorting to other cases. As can be seen in Table 3, several of our other unsuccessful cases provide a less stark contrast, as they might contain a few of the elements (but not all) observed in more successful cases. Before delving into the findings, we first present a brief summary of the two highlighted cases.

Pratham's Teaching at the Right Level (TaRL)

In the early 1990s, the Mumbai Municipal Corporation, in partnership with UNICEF, sought to address the vast discrepancies in performance of primary school children in the city's government-run schools. Conventional wisdom held that low attendance rates were at the root of the problem. Two researchers with deep experience working in the slums of Mumbai, Dr. Madhav Chavan and Farida Lambay, agreed to study the problem. Their research showed that school attendance rates were actually quite high, but that students suffered from a lack of pre-school programs—which had been shown in many contexts to contribute significantly to primary school performance–or remedial programs to address this lack of pre-school preparation. Their subsequent research focused on remedial interventions that could help school children improve their basic literacy and numeracy skills. Chavan and Lambay founded Pratham in 1995, with the mission of "Every child in school and learning well."

In 1998, Pratham began implementing a remedial learning program that identified primary school students who lagged behind and placed them in groups outside of their daily school routine to work with a volunteer tutor ("balsakhi") on basic numeracy and literacy skills. This was the first Pratham program to undergo rigorous evaluation and became the seed of the "Teaching at the Right Level" (TaRL) approach.

Through an iterative process of experimentation and rigorous evaluation, Pratham developed its methodology of grouping students by learning level (rather than by age) and its suite of learning-to-read techniques that are now an integral part of TaRL. Each stage of experimentation focused on a different piece of the puzzle, from the types of exercises and methods that helped individual students learn, to how different types of instructors (independent vs. regular teachers) could deliver the materials effectively, to the logistics of weaving remedial programs into the government schooling system, to mechanisms for efficiently and reliably identifying learning levels in students, to the administrative apparatus needed to provide remedial education at scale. Researchers from MIT's Abdul Latif Jameel Poverty Action Lab (J-PAL) worked closely with Pratham to test the approach in six randomized controlled trials (RCTs) conducted between 2001 and 2014 in six states across India. The results showed that TaRL led to significant learning gains. As of 2019, the TaRL methodology, or variations on it, had been implemented in 21 states and one Union Territory in India, reaching over 16.6 million children, and in over a dozen countries in Africa.

Through its ongoing commitment to the generation and use of evidence, Pratham has consistently influenced the public debate on remedial education, first demonstrating that learning level should be distin-

guished from grade level, then pushing for rigorous and reliable measurement of learning levels, and then partnering with governments to implement evidence-informed solutions.

TARA's Aqua+

The Development Alternatives (DA) Group, founded in India in 1982, is a non-profit organization dedicated to research, development, and distribution of sustainable solutions to address rural poverty. As of 2010, DA Group had a strong track record of developing life-enhancing interventions to assist the poor in rural India. These efforts included developing and/or selling affordable, durable products and services to base of pyramid (BoP) customers, such as solutions for building materials, renewable energy, and natural resource management.

The DA Group consists of three units: Development Alternatives (DA) is a research and development arm, with its own research labs to develop technology and products related to energy, habitat, hygiene, and health. Once a technology is developed and tested, it is transferred to TARA (Technology and Action for Rural Advancement), the incubator of DA Group. TARA incubates microenterprise-based, decentralized models that can sustainably bring products developed by DA to its target BoP markets. Once TARA has designed a business model and distribution strategy for a new product, it typically spins it off as a new, independent, for-profit entity. For several years, DA Group has successfully followed this tri-partite structure, spinning off several independent social enterprises to market its life-enhancing technologies, including TARA Machines (for manufacturing rural construction materials) and TARA Urja (for developing solar micro-girds).

In 2010 DA Group partnered with the Swiss-based Antenna Foundation. Antenna sought to distribute its "WATA" water purification technology to BoP customers, so it partnered with enterprises in multiple countries, including the DA Group in India. Following its traditional approach, DA conducted extensive tests of the underlying WATA technology to ensure its reliability. TARA then took the lead to design a business model to produce, package, and distribute the technology. TARA's fundamental concern was to avoid any possible contamination of the product, so they decided to manufacture small bottles of concentrated droplets, named "Aqua+." Based on this product design, TARA chose a business model that would funnel the product to end-consumers through a network of retail distributors, some already working with TARA. TARAlife was founded in 2014 with the express purpose of marketing the Aqua+ product.

Sales fell well below projections. TARAlife sought to increase sales of Aqua+ by expanding its distribution network to micro-franchisees and through aggressive social marketing campaigns on the health benefits of using Aqua+. TARA was convinced that if households understood the health benefits of Aqua+, they would value the product appropriately and become reliable repeat buyers. But none of TARAlife's adjustments significantly improved sales. As discussed more below, while committed to the rigorous testing of the underlying technology, TARA did not afford comparable scrutiny to all the downstream assumptions built into the Aqua+ model.

Internal and external evidence was available that indicated this kind of product and distribution strategy was not the most effective, as it did not thoroughly consider end-consumers' habits, preferences, and constraints. Other business models for distribution of purified water, which focused on convenience of delivery and ease of use, had proven more successful. By focusing its ongoing efforts on adjusting operational aspects and not on other underlying reasons why customers did not purchase the product, TARA was not able to increase demand for Aqua+. Given the financing structure chosen, which relied on early sales growth to attract future capital, and because of the inherent design challenges noted above, TARA was not able to scale its Aqua+ efforts.

Findings

The Challenges of Evidence in Practice

We identified two fundamental challenges to knowledge coordination in our context. First—and consistent with the general challenges of learning across communities of practice (Carlile 2002))—participants were required to collectively absorb new information, from multiple points of view, and apply it in novel ways. But each community of practice held dramatically different definitions of what constituted timely, reliable, and useful evidence. Each of them also positioned different types of evidence along a hierarchy—the configuration of which varied across groups—where certain types of evidence were deemed superior. Put differently, participants lacked a shared definition and value system regarding knowledge. Since new (types of) knowledge and behaviors often challenged the core ethos of a community of practice, they could also challenge a member's identity, incentives, and sense of authority, presenting a ubiquitous source of friction and resistance. For example, development economists increasingly held randomized controlled trials (RCTs) as the gold standard and only truly reliable form of evidence. In contrast, practitioners often found (rigorous) case studies of successful or failed interventions to be most informative, as experience

had taught them that managerial and implementation practices on the ground mattered most. For implementers, RCTs often were so devoid of context (by design they sought to "control for" it) that they were rendered useless. In the words of an interviewee, *"case studies measure nothing precisely, but RCTs measure, very precisely, nothing,"*—the most important and relevant factors for implementation were left out of the stringent measurement.

A second challenge that amplified the first is that actors not only belonged to different communities of practice but also to different organizations and sectors, each with its own mandates, incentives, and timelines. They lacked a shared goal. As a result, actors within each stakeholder group responded to organizational structures and incentives that had been designed around specific communities of practice and with a relatively insular view of their organization. Academics often faced the imperative of publishing articles in top journals, politicians the constant public scrutiny and periodic elections, and implementers the need to execute programs strictly and according to the proposals linked to each funding stream. Knowledge integration in this context required each stakeholder to venture outside their organization's existing mandate and incentive structures, often deviating from expected performance standards and taking on significant personal and professional risks. For example, to design a pilot or experiment for a new type of government intervention through an interdisciplinary process, academics would need to invest large amounts of time in a process that, in the best case scenario, might open the option to collect unique data and where the path to a publication would be long, winding, and uncertain. A government counterpart in this scenario would need to endorse and help coordinate and implement a pilot with unknown outcomes. It could thus leave a politician exposed to public attacks if the results proved to be disappointing. For both participants, even the smallest of pilots would carry significant downside risk and compete for time and resources with other alternatives that may have clearer and quicker professional benefits.

Together, these two challenges—the lack of a shared definition of evidence and shared goals or incentives—amplified each other and generated levels of uncertainty, instability, and interdependence that far surpassed what organizations were designed to do and what could be reasonably expected of any given actor. Overcoming them required a collective, sustained, and fundamental shift in behaviors and mindsets, as well as high levels of coordination and negotiation across organizations and sectors, as described below.

Effective Integration of Evidence in Practice

The observed challenges were equally present in all our cases, but Table 3 shows that, in some of the cases, actors were able to overcome the hurdles effectively, while in others the barriers prevailed. This was not for a lack of trying or a lower commitment to the integration of evidence. In fact, in all our cases participants were demonstrably committed to evidence, to designing interventions with rigor, to an understanding that the nature of the problem required interdisciplinary coordination, and to the highest standards of accountability. They also all took explicit steps to gather the best evidence, interpret it, and incorporate it into their decisions. The observed differences were thus not in the what, but in the how. Successful and less successful cases exhibited fundamental contrasts in mindsets, processes, and general relationship to knowledge across the design stages that interventions tend to follow (Baldwin et al. 2000; Brown 2019):(1) problem framing, (2) solution framing, (3) process design, (4) project design, (5) prototyping and evaluation, and (6) implementation with its associated evaluation and accountability.

In our less successful cases, actors followed what was often described as the "typical" approach, where they linearly progressed through the design stages. Evidence was conceptualized as abstract, settled knowledge created by researchers, curated by intermediaries who, with some "translation," provided it to policymakers and practitioners. Interventions were designed with a specific set of outputs in mind, and actors assumed that the central challenge would be one of faithful and efficient implementation of the planned solution. While implementation processes were designed to be interdisciplinary, one actor—typically an implementer—took sole responsibility or authority over program design and implementation. Relationships across stakeholders and organizations were typically arms-length and defined by formal channels and contracts that clearly delineated roles and responsibilities. In this view, then, knowledge was assumed to "flow downstream" like a waterfall, trickling down from researchers to final implementers and subcontractors.

Successful cases demonstrated a fundamentally different relationship to knowledge, most evident in a shift from the linear, waterfall transfer of knowledge to a non-linear, relational, participatory, and iterative process where knowledge was constantly constructed, redefined, and enacted through practice. Rather than a pre-established, settled set of abstract concepts, knowledge was treated as an evolving *process*, riddled with uncertainties. And rather than the traditional approach centered around the faithful execution of well-defined tasks within a given timeline and budget, in successful cases actors defined progress and accountability around specific learning objectives.

It generally started with the initial disposition that actors took towards intervention design, as manifested in the early stages of problem framing. Rather than take existing research as settled knowledge, actors in successful cases engaged a wider range of stakeholders and perspectives to jointly identify all that was uncertain or not well-understood about the overall problem space. The starting point, therefore, was an explicit recognition of uncertainty grounded in the acceptance that the problem had not been adequately addressed before. Put differently, rather than fighting against uncertainty and risk as challenges to overcome, participants acknowledged them as an unavoidable starting point that should dictate how they collaborated.

This recognition of uncertainty then translated into a radically different approach that drew out the stages of design. Participants applied a learning orientation throughout the process – from problem and solution framing, to designing processes and projects, before going on to prototyping. Each stage generated new evidence on which subsequent stages built— leading to full-scale implementation that integrated ongoing evaluation and learning. This process, therefore, did not focus on a path to faithful implementation of a well-defined program, but on the design of an inclusive, participatory, and iterative process focused on learning, adaptability, and the progressive resolution of critical uncertainties. This also entailed a fundamentally different definition of accountability.

From this observation, our analysis focused on what allowed actors in some cases to engage in and sustain such unusual and risky extensions. As we will detail below, we found that actors in our successful cases actively and deliberately cultivated a set of stabilizing factors to contain, counter-balance, and support the observed extensions. These stabilizing practices conferred security where the demanded extensions left actors most exposed. While extensions, as counter-normative acts of risk-taking, represented the most notable differences between successful and unsuccessful cases, the less obvious stabilizing factors provided the guardrails, safety nets, and compensatory rewards that protected actors in their exposure, keeping them engaged despite the ongoing demands and constraints of their respective organizations to enable the needs of knowledge coordination.

The Scaffolding Practices that Enabled Evidence Integration

When actors in our successful cases acknowledged the amount and different types of uncertainty that they faced, they also realized that the effective coordination of knowledge called for a collaborative process of learning. This would require new ways of doing their work and engaging with each other, which then presented additional uncertainties regarding coordination and accountability. Once actors accepted col-

lective uncertainty, they naturally embraced unusual degrees of humility, inclusiveness, and adaptability in how they engaged with each other, program design, and implementation. Yet, as they sought a collaborative and integrative approach, they remained formally employed by their respective organizations and identified with their respective communities of practice, each with expectations, incentives, and evaluation metrics that were at best inconsistent with and at worst in contradiction to the type of collaboration required.

Our analysis traced a common set of what we label 'scaffolding practices.' We incorporate scaffolds from previous research as the "processes, structures, and tools that help people solve a problem, carry out a task, or achieve a goal that would be beyond their unassisted efforts" (Wood et al. 1976; Valentine and Edmondson 2015). What is different about scaffolding practices is that they represent a goal-driven, purposeful design that seeks to enable specific behaviors that are otherwise structurally unavailable. At the same time, while they are purposeful and not merely emergent, they are still temporary and informal, only seeking to enable specific behaviors at specific times rather than build a formal or permanent structure (see Valentine and Edmondson 2015; Quintana et al. 2004).

We identified two different sets of scaffolding practices: structural and programmatic. *Structural scaffolding practices* constituted the curation of tangible resources—human, financial, and physical—to provide a broader framework of reliability, stability, and support for *programmatic scaffolding practices* to unfold. Structural scaffolding practices were enacted as overarching, continuous, ongoing sources of support and containment. Programmatic scaffolding practices, in contrast, were patterns of action endogenous to the intervention, enacted in a punctuated manner to organize the work and catalyze a specific type of extension or transformation needed at specific stages of the intervention.

Scaffolding practices played two counterbalancing functions for participants, creating opportunities for effective coordination and learning. The first was an *extending function*, which elicited, informed, and guided the extra-ordinary mindsets and behaviors that marked successful evidence-based practice (as described above and in Table 3). Through the different mechanisms detailed below, the practices enabled and normalized atypical, uncomfortable, and counter-normative behaviors and forms of collaboration.

Scaffolding practices played a second, *stabilizing function*, providing necessary counterweights to allow riskier extensions to take place. They reduced the uncertainty of complex collaborations by providing a common language, coherence, and clarity on the collective goal, as well as validating and protecting individual needs and incentives. The scaffolds provided a sense of relative security and trust in the collec-

tive process, as a counterbalance to the different types of uncertainty surrounding potential (unknown) outcomes. They also gave participants a justification—to their organizations, to their communities of practice—for their engagement with a set of activities, actors, and approaches that fell outside of their usual role and organizational mandates.

Comparing our findings with prior research, we found that while the extending functions provided by scaffolding practices were for the most part well documented, there was little prior identification of the importance or details of the stabilizing functions which, in our cases, played a fundamental role as enablers of the observed extensions. In addition, for those functions that have been well documented, we found that research tends to describe each of them in isolation, whereas in our cases we found that they often needed to operate in conjunction, complementing and counterbalancing each other. Our data pointed to an integrated process, where different scaffolding practices worked together to maintain a tricky balance, providing stabilizing elements to protect each actor in relation to existing expectations and incentive structures, in order to allow for flexibility, risk-taking, and learning. To streamline our findings, Table 4 presents a summary of all the observed scaffolding practices and the extending and stabilizing functions that each provided. It also identifies the connections between our observed practices and prior research, so that subsequent sections can focus solely on the contributions of this paper.

Structural Scaffolding Practices

We identified three structural scaffolding practices that, together, organized the financial, logistical, and organizational resources to provide continuous support for the collaboration process: 1) Convening under a Protective Space, 2) Multivocal Orchestrating, and 3) Securing Adaptive Financing. Table 5 shows the presence of the different structural scaffolds in each of the cases.

Convening under Protective Spaces. A baseline, key challenge shared by actors across our cases was reconciling the demands and mandates of their primary organizations with the needs of the development intervention at hand, which demanded a distinct set of norms, goals, constraints, and hierarchies. It was thus critical to adequately establish and (maintain) a space where participants could periodically come together to advance their collaboration, shielded from their habitual realities. Cases where such spaces were not periodically carved out and maintained faced a dramatic reduction in the extent to which participants were willing and able to sustain engagement with an intervention. In such cases, actors with higher formal authority over the process typically (sometimes unwittingly) imposed their perspectives

and timelines. Others, in response, were only willing or able to engage as far as their existing professional and organizational constraints allowed—partially, sporadically, and insufficiently.

As shown in Table 4, much of what we observed regarding Convening under Protective Spaces was consistent with existing research—these spaces demarcated a joint time and space apart from the occupational and organizational structures that each participant resided in. Yet, our research unveiled important nuances surrounding the dynamic need to manage and reconfigure such spaces over time.

On a practical level, convening others under a protective space responded to an observed need. If an intervention required diverse and sometimes conflicting stakeholders and perspectives, then a critical starting point would be a context that enabled a shared understanding of the core tenets of the problem and the process to address it. At the same time, while the need for convening under a protective space was consistent, the specific needs of participants changed significantly over time, so the actual type of protection, of convening, and of space also changed to adapt to the evolving nature and needs of the intervention.

For example, in the early stages of a collaboration it was often most critical to establish a shared understanding, a shared set of goals, rules, and norms—*who* would be involved, *why* each participant was necessary, *what* each of them brought to the table, and *how* they would all interact. In our successful cases, this was typically established through a structured conference or official gathering, with active facilitation directed at jointly answering key questions to settle on a joint approach.

Pratham has learned to always convene diverse stakeholders—including teachers, school administrators, local government officials, national policy makers, and academic researchers—in an inclusive and collaborative approach to design, implement, and evaluate its initiatives. Through experience, Pratham has learned to replicate the same general framework every time it enters a new region, pilots new initiatives, modifies core elements of its program, or undertakes a major revision to its methodology. It all starts with a structured convening, as described by a participant:

> Typically it's a workshop or a short half-day session, where we ensure that we have all the right stakeholders, and we're not looking at just one [government] secretary but we are also talking to his other peers, because he or she may not be in the seat over the next six months. We ensure that it's a broader discussion, an evidence-sharing session.

During these initial sessions, Pratham's approach ensures that all voices are solicited and heard, including those skeptical of TaRL. Through active facilitation, there is a set of questions that participants grapple with to establish a shared baseline. Pratham will often end the sessions with a joint call to action, a

proposed path forward, an agreement on the norms that will guide the collaboration, and the general roles that different actors will play.

Important as that initial convening may be, Pratham (like conveners in our other successful cases) has also learned that, once a joint path is established, participants who have subscribed to it continue to need protection but of a different type. A temporal shield to experiment with alternative norms and identities is critical, for example, during the problem framing and process design stages. But once actors have subscribed to an approach and move to implementation, they are most in need of a formal justification to suspend and depart from their traditional role expectations. In response, Pratham typically convenes actors to a new protective space, in the form of a working group or task force that is both more formalized and explicitly bounded in goals and time. The protection—as justification and excuse—is typically embodied in a formal invitation for participants to join the working group. As the central convener, Pratham strategically facilitates the type of protection that is most useful for each actor. For example, for academics it may be a letter of intent for a research collaboration with a government entity. For school teachers, in contrast, it may be a formal joint invitation from the ministry of education and Pratham, sent to them and their supervisor, to participate in the task force—describing the importance of the work, the time and effort the project will require, and the frequency with which teachers will require time away. For all participants, the structure of the task force becomes a "permission slip" to wonder outside of their traditional role expectations.

The needs for the protective space might change again when the actors need to assess, share, and adapt to key learnings at different moments of the collaboration, so Pratham typically convenes a new short conference, again featuring diverse stakeholders, to share experiences, digest lessons, and agree on a new path forward, and so on.

In summary, while there was a consistent need for protective spaces throughout a collaboration, there was also a need for an active assessment of what type of convening and what type of protection actors needed most at different stages of the collaboration. At certain key moments—most notably at the beginning of a collaboration and when agreements needed to be renegotiated to adapt to learning—the most critical aspect of a convening was to actively enable extensions in mindsets and behaviors, such as establishing new shared identities, norms, and ways of relating. But once those extensions were activated, participants were in alignment and eager to work together. The most important function of a convening space then became to ensure a protected *stability*, a sense of security as actors felt shielded from outside pressures

and felt that their extensions and efforts were also demarcated in time and scope. This helped actors sustain those extensions in the face of the continuous demands and expectations of their communities of practice. Convening under protective spaces was thus a dynamic, structured, and managed practice.

In contrast, not providing an adequate and ongoing shield for participants from external demands and expectations prevented even the most committed actors from participating in a sustained and sufficient fashion. Interventions that lacked dynamic, adaptive convening under protective spaces often resulted in a form of breakdown, most typically in seeking cooperation from critical stakeholders only once a crisis emerged. This often resulted in frictions, resistance, or the discovery that it was "too little, too late."

For example, TARA's parent organization, DA Group, was split into discrete units, with Development Alternatives carrying out the basic scientific research, TARA designing the product and business model, and TARAlife assigned responsibility for distribution of the Aqua+ product. Within the DA Group, there was no structured opportunity for all those involved in Aqua+ to share perspectives or to solicit inputs from external stakeholders, such as its NGO partners and retail franchisees. While the scientists within Development Alternatives conferred on how best to test for and validate the product's efficacy, that kind of convening focused entirely on the underlying basic science, and did not identify the need to explore how best to design and distribute what later became Aqua+. Interestingly, TARA itself participated in an international network of country-level partners under the Antenna umbrella, with periodic convenings to share best practices and lessons learned. That convening, however, sought to contrast differing business models and delivery mechanisms across countries using the Antenna technology, so TARA discounted its relevance to the business model it had already chosen in India, grounded in their numerous prior successes.

Multivocal Orchestrating. Convening under protective spaces was necessary to bring together a diversity of voices and perspectives and form a cohesive path forward. But such diversity also brought with it complexity, contention, and frequent misalignment. Within established organizations and work groups, hierarchy and formal processes can provide clarity of direction, define priorities, and settle disputes. In our settings, without a clear and shared mechanism to determine how different perspectives would be organized and prioritized to make decisions, some interventions quickly devolved into a gridlock of competing priorities or the imposition of a dominant but limited perspective. What allowed some interventions to coordinate effectively was the practice of multivocal orchestration.

Once again, as reflected in Table 4, the extending functions of orchestration—mediation, conflict resolu-

tion, active translation—was consistent with prior research. This includes the observation that it was, in all of our successful cases, performed by either (a) a dominant actor in the program with an interdisciplinary background that enabled them to understand, interpret, translate, and fairly represent multiple perspectives; and/or (b) a neutral mediator who was recognized as having no personal or organizational agenda beyond the success of the collaboration. Much of orchestration entailed anticipating, identifying, and clearing disagreements through mediation. Given this role as brokers and mediators, it was critical for orchestrators to be perceived by all stakeholders as legitimate and neutral. As participants recognized the amount of uncertainty surrounding the path and destination of their collaboration, trust in the neutrality and legitimacy of the orchestrator translated to a trust that, even if the outcome was uncertain, it would be fair and would honor the needs and commitments of each participant.

Both Pratham and its research partner, J-PAL, developed extensive experience providing multivocal orchestration in the TaRL programs. Pratham established itself as an independent organization committed to designing effective education programs and then disseminating rigorous evidence through working groups, task forces, conferences, and publications. Its inclusive approach—integrating program delivery with both research and public policy—has led it to become a trusted convener and interlocutor in tackling complex issues of remedial education. Pratham's capacity to play this role has been enhanced by the fact that its founders have straddled multiple sectors throughout their careers, including stints in academia, public service, and business. Their partnership with J-PAL has further strengthened this perception of neutrality and legitimacy:

> The role of J-PAL was critical (...) an external entity endorsing a program based on objective evidence (...) They were an impartial organization that understood and had done the research and that helped us coordinate around more objective factors, like existing evidence (...) Just the brand of J-PAL and the universities of its affiliates, that puts people at ease immediately. If you say, "Oh, MIT, Okay," it is a name they have trust in. - *TaRL Researcher*

We also found that orchestration served two instrumental functions in our successful cases. First, we found that some of the most important orchestration happened behind the scenes, as a proactive and frequent process of engaging with each stakeholder to assess their experience, identify their needs and concerns at each stage, and ensure that a corresponding set of scaffolding practices (more on this below) were put in place to provide stability and ensure that participants could focus on "the work." In their TaRL projects, for example, J-PAL met with each participant periodically, both formally (e.g. programmed progress meetings) and informally to keep very close track of progress, concerns, and challenges. It is through such informal contact, for example, that they identified the need to formally invite teachers to a task force so they could justify and protect their time investment (from their supervisors and peers),

which meant that the invitation had to come from the ministry of education and Pratham collective.

The second function was the more prosaic yet instrumental and consuming labor of day-to-day project management, coordination of logistics, and administration of collective resources. For example, we have described the importance of periodically convening participants under a protective space. Even the simplest convening—and some, like conferences, followed complex formats—requires significant organization and logistics; from identifying and inviting the right participants, to securing venues, setting agendas, facilitating the right type of interactions, providing nourishment and logistical support, keeping track of activities and agreements, and following up on next steps. In some of our cases, actors with more experience in orchestration like Pratham developed internal capabilities to periodically organize and manage such events: "It's not rocket science, but doing it well can be very important (...) If you do it right, nobody notices, but if you do it wrong, it can ruin relationships." Pratham and J-PAL have thus learned how to facilitate the integration of diverse stakeholders to bring them along in a managed process of collaboration:

> We invited all the teachers of the Delhi government to a workshop so they had a real voice in the design (...) "look at this textbook in a group and tell us how appropriate this is for the child in your class. (...) Can you create something more simple with which you can connect with your child?" The teachers started giving feedback because we opened the debate. Before, it had been closed: "Here is the syllabus, here is the content, you have to deliver that." And we managed the feedback system and follow-up.

Orchestration, then, was not only about mediating and facilitating extensions between actors—important as that may have been—but also about the managerial and structuring practices through which participants found the necessary stability, continuity, and clarity to engage in the extending behaviors required by the work. In all of our cases—successful and not—there were effective, experienced project managers. The difference, however, was that in unsuccessful cases project management focused only on the achievement of specified goals in time and under budget, through pre-defined tasks and contracts. In successful cases, in contrast, project management was one of several functions of multivocal orchestration. And project management was carried out *through* multivocal orchestration, including the definition of goals, timelines, and budgets, which were frequently and collectively redefined.

Cases that lacked multivocal orchestration often followed top-down approaches led by an authority within the organization leading (or formally charged with) the intervention. Without facilitated orchestration, intervention designers did not identify, understand, or trust alternative and necessary perspectives. There was no mechanism to identify the needs of other participants to match them with the corresponding supports. The absence of multivocal orchestration, for example, often correlated with the failure to convene

participants under protective spaces, as both stemmed from the under-appreciation of the complexity of the challenges or the need for different types of expertise. The result privileged certain, central voices, while excluding other critical ones.[4]

On one level, the DA Group is multivocal by design, covering basic research, intervention design and experimentation, and then spinoff of independent organizations for each solution. At the same time, the core teams engaged in Aqua+ at each stage were part of separate organizational units, and no effort was made to integrate their varied perspectives. Rather than facilitate a shared process, there was a linear hand-off between them. This had actually served the DA Group well before, in the development of solutions that had less uncertainty and required less inter-disciplinary learning. In the case of Aqua+, however, the approach missed the perspectives from those on the front lines of distribution, both within and beyond the Aqua+ network, including distributors and store owners. As described by TARA staff:

> Different teams, different people handle different aspects of the product, where everyone feels that they're working for promoting Aqua+ (...) I think in bits and pieces everyone is working on some aspect of the product, but there is no integrating structure.

Adaptive Financing. While prior research would have predicted the importance of Convening under a Protective Space and Multivocal Orchestrating, it had much less to say on financing practices. Yet, for the diverse stakeholders that came together in our cases, the approach to financing was among the most important for the effective integration of evidence. First, because there was no shared organization and budget, so financing committed to their collaboration was a necessary glue to bring stakeholders together. Second, funding streams bring resources but also restrictions. Those who allocate funds and oversee their deployment must remain accountable to the taxpayers or private donors who entrusted them. The top priority, reflected in rules, values, and practices of funders, is to ensure the transparent, judicious, and ethical use of resources. Accordingly, financing agreements typically prescribe budgets, activities, timelines, deliverables, and audits as mechanisms for accountability. But this is in tension—and can be in contradiction—with the investment required for discovery and the flexibility required to innovate or adapt to new learning.

Financing can thus be highly constraining to the effective generation and integration of evidence, as it can commit stakeholders to a sequence of actions or a set of outcomes that new information may reveal to be unrealistic or less appropriate than anticipated. At the same time, dependence on critical financing often leads actors to accept strict, narrowly-defined contractual and budgetary terms that fail to incorporate the complexity, uncertainty, and interdependencies of an intervention. Once committed to a financing

contract, participants are bound by its budgetary rules, including on which adjustments can be made, at what times, and how.

> The fact that we adjust our approach as we learn every year makes us a bit unattractive to corporate donors because corporations don't understand the objective of doing that. They are in the mindset of 'why should we change the way we are doing something if it is working?' - *TaRL Implementer*

In our successful cases we identified flexible, adaptive financing practices that allowed for a more fluid relationship between partners. Instead of starting from a defined set of parameters; activities, timelines, and outcomes were gradually shaped through learning at each stage of the intervention. Adaptive financing provided the stabilizing forces of a reliable path to resources and measures of accountability. But these were balanced by the extensions facilitated by flexibility in the way the resources could be used in time.

Actors secured adaptive financing in one of two ways. In some cases, they negotiated open-ended financing from the outset, explicitly avoiding definitions of specific activities or outcomes, focusing instead on a broad programmatic vision that emphasized learning and allowed for a range of possible learning-based trajectories. In others, while certain players, activities, and outcomes were defined at the outset, actors established a flexible financing trajectory, with built-in mechanisms to adapt parameters dynamically as learning arose. Specifics aside, what was common across our successful cases was that, once participants identified key uncertainties upfront, they also recognized that financing must be flexible to allow for learning. This then shaped how they negotiated for adaptive financing structures. Adaptive financing could be secured from a single funder, or it could be patched together as a path of financing, mapped for adaptation, from different funders.

For example, to develop its "Read India" program, Pratham negotiated three successive rounds of funding with the Hewlett Foundation. Pratham preferred funding structured as a sequence of rounds, each focused on a single cycle of experimentation, evaluation, reassessment, and a pivot to the next cycle. In 2007, Hewlett awarded a grant for the first iteration of the program, with a focus on experimentation at scale. It was implemented directly by Pratham staff with government partners and was accompanied by rigorous evaluations. In the first two years, Pratham scaled quickly, reaching 33 million children. In 2009 the Indian Government passed the "Right to Education Act," which replaced year-end examinations with "Continuous and Comprehensive Evaluations" and required children to move to the next grade level through ninth grade according to their age irrespective of their learning—the opposite of TaRL's philosophy. In this context and despite its proven success, Pratham negotiated a second grant with the

Hewlett Foundation that actually *scaled back* Read India. In this second phase, Pratham took a step back and implemented its model in a few strategically located blocks throughout India to "illustrate to key players in education that rapid and cost-effective improvement in learning are possible using Pratham's methodology and materials." This translated to a reduction from operating in 300,000 villages to only 25,000. Rather than focus on scale—as the original trajectory suggested—Pratham sought to build deeper relationships with local decision-makers and communities to "bullet proof" Pratham's evidence-based approach focused on learning from political cycles. Lessons from this second cycle informed a third, also Hewlett-funded, phase of Read India (2013-2016) that sought to scale TaRL both through government schools and direct Pratham delivery, according to the local context. As of 2019, the government partnership model had reached 15 million children across 15 states in India, while the volunteer-led model reached 1M students across 18 states. Pratham leveraged the results of the accompanying RCTs to secure hundreds of district-level government partnerships.

To build additional adaptability to its financing portfolio, Pratham also routinely secures funding explicitly detached from program implementation to build organizational capacity, nurture its data-driven culture, and to test and refine its approach in smaller experiments. One key element is finding funding partners with a shared commitment to learning: "it was more that we have a shared goal and we're all trying to discover the best way to reach it, than in a traditional directive 'we're the funders, it's our money, and we think this is what you should do, so do it'." Pratham has learned that a focus on adaptability requires an unconventional approach to financing that seeks longer-term relationships with funders, sometimes asks for *less* money than is offered, is structured around stages of learning, seeks unusual sources of funding that are nonetheless aligned with learning objectives, and turns down funding that seeks to impose traditionally rigid terms.

In contrast, interventions that established financing structures aligned with more traditional implementation objectives, faced enormous challenges when trying to adapt to new learning. In the Mexican case of PPE, for example, the federal government's political timeline dominated the design and implementation process. The policy was designed to meed a set of specific outcomes at specific times, but it failed to identify key uncertainties, creating rigid budgetary restrictions. Even as evidence emerged that objectives were not being met for reasons uncovered during the process, and even as key stakeholders sought to help provide necessary adjustments, the constraints imposed by federal budgeting laws limited potential changes:

> The federal evaluation agency sent a basket of recommendations, some of which were addressable within the scope of PPE and its published operating rules. But the most important ones fell outside that scope, as they implied a different approach. We looked at those recommendations seriously and we spent a lot of time looking for ways to address them, but ... they went beyond what we are legally allowed to do, which is to follow the law in the use of federal resources and program implementation. All we could do is tweak it at the margins, which we knew was not enough. - *PPE Implementer*

To be effective, adaptive financing practices needed more than flexibility. Antenna provided flexible financing terms to each of its country partners, including Aqua+, consistent with its aim to learn from the different approaches to the distribution of the WATA technology across geographies. However, Antenna adopted a portfolio approach to compare and learn across settings, with no specific learning goals—or identification of key uncertainties—within each investment. Antenna's limited funding thus lacked a counter-balancing rigor for its partners to identify learning needs, document pilot results, and disseminate lessons. Perhaps more critically, TARA's financing strategy assumed that the key challenge would be a faithful implementation of its design. It was thus premised on raising funds for scale-up based on the successful—and thus profitable—execution of its business model. The lack of success in the initial implementation—not because of mismanagement, but because of the uncertainties that were uncovered—also meant that the strategy for subsequent funding had nothing to build on, or a cushion to absorb unexpected uncertainty. Their funding strategy, however, matched their prior experience:

> It's a larger problem. Development organizations and social businesses underestimate the amount of funding required for data collection and analysis that would lead to robust 'evidence building.' Donors and investors want us to 'implement,' ... but are extremely wary to support evidence-building capacity within social businesses and have not been willing to embed monitoring, learning, and evaluation partners together in project teams. - *Aqua+ Implementer*

Programmatic Scaffolding Practices

Structural scaffolding practices set baseline conditions, provided basic stabilizing elements, and secured appropriate resources to enable novel forms of collaboration—extensions—between participants. They did not, however, provide answers to the basic questions of organizing regarding how to distribute work and coordinate effort—or to what end. In our successful cases, we observed five distinct and complementary programmatic scaffolding practices that, together, established clarity and coherence of direction; agreements on guiding principles, collective values, and expectations; and an organizing framework to make decisions. These scaffolding practices emerged endogenously as common, effective responses to the challenges that the lack of a shared organization created for participants at each stage of the work.

We identified five programmatic scaffolding practices: 1) Negotiating a Shared Problem, 2) Establishing a Learning Orientation, 3) Distributing Expertise, 4) Pricing Currencies of Exchange, and 5) Mapping

Learning Pitches. Similar to their structural counterparts, and as reflected in Table 4, some of the functions provided by the programmatic scaffolding practices have been documented before. We thus focus on the aspects and functions that emerged as contributions of this paper.

Negotiating a Shared Problem Stakeholders came—or were brought—together based on a trajectory of work around a similar set of issues or populations. But they belonged to a variety of organizations with disparate and sometimes conflicting mandates, approaches, and visions. Beyond overlaps in their trajectory, there was no guiding principle for how, or to what vision, collaboration should happen. And no shared hierarchy to bring coherence and define priorities.

To find coherence and establish a shared vision amidst this complexity, actors in our successful cases actively engaged in Negotiating a Shared Problem. This entailed an extensive, structured, and facilitated process that guided them to jointly define a focal issue faced by a particular population that they were all invested in. As part of this first scaffolding practice, participants were led to identify a set of values that they shared, despite their differences in how they had traditionally approached the issue space. In all of our successful cases, these shared values were grounded in the participants' shared commitment to the target population. The shared values were also explicitly identified as the beacons that would guide decisions and help settle disagreements. And from these values, participants negotiated a shared problem that was explicitly framed as the vision for their collaboration, with its accompanying priorities.

In all of our successful cases, this negotiation was initiated, facilitated, and structured by the central convening actor. It created a framework for actors to grapple with an issue space that was deeply known and shared, but from different perspectives. Because they all found the problem relevant and compelling, and because each had typically approached it from a different perspective, the process revealed that the solution was not known and would need to be discovered in collaboration: clearly, all their (isolated) attempts at solving it had not succeeded. Negotiating a Shared Problem thus provided stability by clarifying a shared goal, identifying a shared language, and articulating the shared values that would drive decisions and help settle disagreements. This stability then provided the space for risk-taking and collaboration.

As noted earlier, Pratham launched each new adaptation or expansion of TaRL by convening the relevant stakeholders, often through multidisciplinary tasks force, to conduct a structured joint assessment of their area of interest to arrive at a shared understanding of the specific problems they were trying to address, as well as the specific challenges and opportunities provided by the context. With that rigorous assessment,

Pratham then integrated all relevant stakeholders and facilitated a thorough discussion of the problem, to arrive at a joint diagnosis and problem statement. Throughout the process, Pratham helped actors identify their joint commitment to children and their learning as a stronger force than their disagreements. Pratham also facilitated a structured process to help actors articulate a set of principles that they all shared, always rooted in learning. As one staffer put it:

> In new locations, especially in rural areas, it made sense to try and figure out if children needed our help with learning to read or for basic arithmetic. We began to work with local villagers and communities to create what we called 'village report cards.' Every child in the village was asked if they were in school and then given a simple story to read and some basic arithmetic tasks. This was done hamlet by hamlet and neighborhood by neighborhood. Neighbors, friends, siblings, all helped. With these data, the assessment was accompanied by a lot of heated discussion and debate [among stakeholders]. Finally, the conversation would move towards what we wanted to achieve.

In contrast, once TARA had tested the effectiveness of WATA, they made relatively fixed choices to manufacture the product and distribute it in the form of small bottles with a drop dispenser. This, they believed, would ensure fidelity and reliability based on their water purification research. Since product distribution and household take-up were not defined as part of the fundamental problem, they were not accorded the same degree of exploration, consultation, and testing as had been the case for the purification technology. In the face of disappointing sales and negative customer feedback, TARA expanded its social marketing and education efforts—"if individuals did not change their consumption habits, it must be because they did not understand the benefits of the product." But TARA did not reconsider the underlying product design and distribution approach. Rather than negotiate a shared problem with key constituents across the value chain—including the small store owners who would need to carry the product to reach end customers—TARA derived an internal definition of the problem, based on its research and its prior successful experiences, and assumed that it would work for this new set of stakeholders. Looking back at the difficulties encountered in distribution, one TARA employee noted:

> I think we didn't do enough at the customer end ... our belief at that point in time was that if we got our equation right with the distribution channel partner and provided enough of a margin, then they would do the selling on their own. That didn't happen, except in a few cases.

Establishing a Learning-based Orientation. In development work, effectiveness is typically defined as the delivery of expected results within a set of prescribed outputs, including budgetary thresholds and implementation timelines. Professional roles (notably funders, implementers, and policymakers) require a results-based mindset, with an emphasis on accountability.

Despite their traditional roles, actors in our successful cases shifted from their focus on implementation and maintained an orientation towards learning throughout the interventions' cycles of design, piloting,

and implementation. This orientation did not emerge spontaneously, but was actively cultivated from the outset as the second, facilitated scaffolding practice.

Negotiating a Shared Problem had allowed actors to agree that, while they collectively knew much, they did not know enough. Accordingly, as a complement to that first practice, organizers led participants to recognize that, despite their individual and collective work, they had failed to make satisfactory progress. The priority should thus be to *discover* what could work, rather than simply modify programs that had not delivered on their promise. Participants were led to identify what was well known, where there were gaps in existing knowledge, and the types and levels of uncertainty that remained—regarding, for example, basic intervention design, mechanisms of implementation or adaptation, or broad scale-up.

Establishing a Learning Orientation released participants from their typical focus on implementation outputs, extending them to accept (rigorous) collective learning as an alternative measure of success. It also allowed for exploration, where "failure" was re-framed as a necessary part of learning. An orientation towards learning not only helped participants identify uncertainties, but also revealed the need for a different approach to collaboration, which in turn uncovered new challenges and the need for subsequent, corresponding scaffolding practices. In that sense, the practice of establishing a learning orientation was the entry point to an alternative trajectory of collaboration.

Pratham's general approach has always been to improve educational outcomes, following the best available evidence. It has cultivated longstanding partnerships with research and academic institutions, including J-PAL. These learning partnerships constitute the core for each new initiative, to which other critical stakeholders are invited according to the context's needs. Through the convenings described earlier, participants identify the collective state of knowledge, the key pillars of what is known and should inform the new collaboration, as well as the gaps in knowledge that the new partnership can help discover. Pratham's research partners may present the results of their latest studies, practitioners may discuss the challenges they faced implementing the work, and policymakers will discuss their priorities and political landscape. It was through these discussions—including about changes in the policy landscape—that Pratham and partners determined that, despite their operational and funding success, phase two of Read India should shift its focus of learning, from questions of rapid "top-down" scale-up, to questions of how to systematically establish local partnerships and ensure "bottom-up" ownership of TaRL.

Through experience, Pratham has learned to routinely establish a learning orientation in every partnership. This is reflected in formal organizational practices, such as its Measurement, Monitoring, and

Evaluation (MME) unit, which grew from 12 members in 2013 to 80 by 2016. MME was founded on the premise that people are typically held accountable to whatever is measured by their organizations, which creates powerful incentives. Yet, learning is rarely measured or used as a standard for accountability. The team has thus learned to present data strategically to different audiences to generate incentives for learning. For example, Pratham initially experienced resistance to a tool designed to help teachers track individual children. After Pratham's MME engaged selected groups of teachers, presenting use cases that demonstrated how they could use the data to do their jobs better (and more easily), demand for the tool vastly expanded, leading to widespread implementation. In addition to such "pull" mechanisms to induce bottom-up participation in the collection and use of data, Pratham has also learned to "push" incentives. For example, to spur competition between two critical districts, MME would send key local policymakers—and their external constituencies—a periodic report comparing learning levels and rates of progress across districts.

In contrast, Aqua+ developed through three phases with very different learning orientations. During the first phase, the DA Group framed its research as a (yes/no) validation of Antenna's claims about the effectiveness of WATA. The DA Group, historically committed to rigor, conducted a series of scientific experiments to ensure that WATA met the highest standards for reliability and efficacy. Once the technical threshold was met, the learning cycle was deemed complete and WATA was transferred to TARA.

The second phase focused on developing a product that would ensure the integrity of the technology. TARA's priority, informed by phase one's rigorous testing, was product effectiveness and quality, which relied on strict levels of solution purity. But TARA did not pay comparable attention to potential downstream challenges, such as distribution and marketing, and failed to identify nested and downstream uncertainties beyond the effectiveness of the technology. This narrowing of the learning orientation proved to be the Achilles' heel of Aqua+. TARA assumed that if the technology worked, the product was carefully produced and made available, and its health benefits were clearly communicated, target households would purchase and use it. Finally, during the third phase TARA sought to define a sustainable business model that could deliver the product, with technological fidelity, to consumers. Even though TARA recognized that consumers would need to develop new habits, the marketing strategy centered on education, and developed from a fixed conception of the initial product design: the assumption that when communities were educated on the risks of drinking from polluted sources and the resounding benefits of using Aqua+ at a minimal cost, demand for the product would grow:

> ...the marketing strategy was almost, I wouldn't say frozen, but it was finalized. We would move ahead with finding (distribution) partners (...) This is a consumer product, our job was to get the product and message to its target market in rural India.

Leveraging Distributed Expertise A learning orientation helped actors in successful cases identify key nested uncertainties and knowledge gaps—critical questions that their collaboration would need better answers for. But, as described before, stakeholders had radically different definitions of what constituted timely, reliable, and useful evidence; or what would count as a better answer.

Yet, despite their evident disparities, participants in successful cases described an alternative collective agreement, where knowledge was no longer treated as an abstract entity residing in a hierarchy dominated by the rigor of RCTs. Rather, they described *bodies* of complementary evidence, where different types of evidence, emerging from different types of methodologies, and reflecting different types of expertise, were best suited for different gaps in their collective knowledge. ((Participants still discussed the importance of rigor, but they also acknowledged that it could mean different things across approaches, and that these should be determined by what would be most relevant, timely, and useful to each aspect of the problem.))

This agreement was achieved through the scaffolding practice of Leveraging Distributed Expertise. It entailed an active, facilitated process whereby, based on the uncertainties and gaps they had previously identified, participants came to jointly agree on which actors had the most relevant knowledge and expertise for which elements of the problem. In our successful cases, this facilitation first took place during an early convening of participants (in some cases in the first structured convening). Facilitators actively solicited and incorporated every perspective, leading the group to match areas of uncertainty with the actors who were best positioned to help address each one. The scaffolding practice then extended to later parts of the process through multidisciplinary steering committees and decision structures. The observed result—or behavioral extension—was that participants suspended their conventional hierarchies of evidence and authority as they acknowledged their interdependence. But this extension was actually grounded in the stabilizing conviction that their respective expertise and authority would be appropriately incorporated into the most relevant parts of the process.

Pratham has established the practice of creating a formal steering committee for each new intervention, where not only every stakeholder is guaranteed voice before each decision, but also where tacit and explicit norms ensure that the actors closest to each element of the methodology will have more influence (e.g. teachers and union representatives on how to adapt the TaRL model to local practices, education ministry staff on government financing and rollout, academics on the different types of intervention design

that would allow for the most rigorous learning for each specific question). As a result, Pratham and its partners routinely draw on a variety of research methods and approaches, including RCTs, theoretical papers, qualitative case studies, large-scale data analyses, and local community surveys. In our interviews, researchers, policymakers, and implementers agreed that, while RCTs have played an important role in legitimizinig TaRL, especially through the international academic community, they are one "slice" among many types of evidence required for program design and implementation. Other types of learning have proven equally useful at different stages of the process. For example, Pratham launched a national Annual Status of Education Report (ASER, discussed further below), which allowed for a different type of learning and conversation:

> We need ways to engage people so that learning can become more visible and understandable... show activities in which ordinary parents and family members, neighbors, and youth can engage. The purpose of the assessment activity in the village report card was to enable a broad cross-section of people to participate, so that they can understand the situation and take action.

In contrast, once TARA (through its parent, DA Group) had validated the WATA technology, it did not seek other perspectives or evidence to inform its process. TARA subcontracted a market research study with Delhi doctors, chemists, and shop keepers to explore potential distribution strategies. On the demand side, the study interviewed potential (urban) customers who said they would prefer to buy the proposed product from a chemist or grocery store. They were not, however, asked about the underlying appeal of the product itself (Aqua+ drops in a small bottle), which was taken as given, or how the need to use the product consistently and purchase it routinely from the vendor would fit with their consumption and purchasing habits. Given that TARA had not identified downstream uncertainties, including on how to establish relationships with a large network of small-store owners or on ways to change the purchasing and consumption habits of its target households, it did not incorporate actors who could bring those perspectives into its design, implementation, or evaluation processes. TARA framed the problem entirely as a product + operations challenge, which is its core expertise, so it assumed that all relevant expertise was contained in-house.

The focus on operational issues, grounded in previous successes, led TARA to neglect the drivers, motivations, and habits of its target households. The staff considered a variety of potential distribution and business models, selected one, and made adjustments as they implemented it over time. But there was no comparable effort on the retail or customer behavior sides, which academic research and the experience of other Antenna partners had shown to be of critical importance:

> Everyone still talks about solutions at the distribution level and nobody is really looking at the needs and

> habits of the consumer, building solutions where it's sort of at their doorstep. We sort of shifted away from the evidence that was there in terms of studies (demonstrating the importance of behavioral factors) and relied only on the evidence that we had built over the years in our own experience. - *TARA Implementer*

The initial framing of required evidence meant that TARA did not develop a systematic way to gather, interpret, and evaluate downstream evidence, not only from critical stakeholders but also from its own operations. Consistent with its chosen business model, TARA's distribution strategy focused on finding "hub partners" that could take ownership over distribution in many, often remote, locations. This was efficient from TARA's perspective, but it also meant that TARA remained several steps removed from the field. The organization also lacked the funding and capacity to reliably and routinely collect household data—they did not plan for it.

> We didn't do a lot of monitoring, I think. I mean, we did some surveys and we're still waiting for the analysis, but I have not had information about that [households]. - *TARA implementer*

TARA also discounted external evidence it was exposed to, as it was collected through approaches that, on the surface, seemed inconsistent with TARA's chosen strategy. For example, TARA learned about Spring Health's market success delivering large jugs of pre-purified water from its own network of kiosks. While this model offered critical insights on how distribution was framed to fit within user habits, motivations, and needs, TARA overlooked these insights, as they were embedded within a somewhat different business model. As one employee said, "...the mandate has always been to find the most efficient delivery process for the product [as designed]."

Pricing Currencies of Exchange . The scaffolding practices described so far provided stabilizing mechanisms that helped actors relax their assumptions, goals, and habits to allow for openness, flexibility, and attunement to others. Yet, every participant was still beholden to the incentives, mandates, and constraints of their respective organizations. Regardless of their disposition and orientation, actors still had full-time jobs to deliver on.

To carve and incentive space for collaboration, actors in successful cases undertook extensive negotiations to "price" each other's "currency of exchange:" the assets, knowledge, and capabilities they brought and the needs, incentives, and constraints that each faced. This allowed for a genuine investigation and allocation of individual interests and constraints in relationship to the shared problem. This scaffolding practice collectively clarified what each party had to offer and what each needed to get in return to remain committed to the work. The same diversity that made collaboration challenging made it valuable, as actors possessed different types of complementary currencies. For example, researchers could offer the rigor of their questions and methods and the legitimacy of their institutional affiliations in exchange for

access to government data or to novel settings that would enhance their research. Policymakers could offer access to invaluable data in exchange for the legitimacy of a neutral, external academic evaluation. The practice generally was first deployed when participants moved from the identification of uncertainties to the design of the process to address them. It was redeployed as needs or constraints evolved.

Specifically, we identified six attributes—organizational incentives, unique resources, non-negotiable needs, evidence priorities, time cycles, and target audiences—that were explicitly traded and shaped the currencies of exchange across stakeholders. Table 7 provides additional details. For example, while researchers often measured time according to publication cycles, tenure, and grant timelines; practitioners structured their work around funder requirements and end-client needs. Policymakers, in turn, responded to electoral cycles and the "tyranny of unexpected events."

Pricing currencies of exchange often required a laborious and time-consuming process, given the multiplicity of stakeholders involved. And to be most effective it required the synchronous presence and participation of all, to create common knowledge of their respective resources and constraints. But it created a critical stabilizing force, as participants could trust that, as they committed time and resources to the collaboration, their individual needs and incentives would be honored. As with other scaffolding practices, pricing currencies of exchange required effective, multivocal orchestration to ensure that, keeping the shared problem at the center, every actor's needs and incentives were fairly identified and balanced, without one stakeholder's currency dominating. For example, accurately pricing policymakers' currencies of exchange could be critical to secure resources for a program over time; however, allowing this currency to dominate could lead to abbreviated timelines (synchronized with electoral or media cycles), making it impossible to collect data with the rigor required to generate reliable evidence.

Through its experience, Pratham has become extremely sensitive to the trade-offs between program design, rigorous research, and careful implementation. It has also learned to uncover those trade-offs early to facilitate compromise and exchange. This starts with their main partner. J-PAL and Pratham have learned to identify, express, and support each other's needs throughout their partnership. J-PAL can test large-scale pedagogical interventions through rigorous RCTs and publish the results, while Pratham can build its understanding and international credibility by collaborating with prestigious researchers. This also extends to other partners. Pratham has learned to balance political needs and cycles, for example, giving constant public recognition to government partners—"we never call it our program, we always let the political people name it and they are the owners." Implementation details and rhythms, in

contrast, are dictated by the constraints, needs, and cycles of teachers and their schools. And Pratham helps protect these rhythms from the pressures for quick results and premature announcements from politicians.

In 2010, the state government of Haryana asked J-PAL to help build the government's internal capacity to conduct RCTs—in line with an international trend to build "government labs." J-PAL brokered a partnership with Pratham to bring TaRL to the state. J-PAL saw it as an opportunity to assess how new federal policies affected TaRL; Pratham hoped to demonstrate that TaRL could scale effectively from state governments, and the local government was eager to legitimize their evaluation efforts through their partnership with prestigious academics. Beyond these complementary interests, actors had to learn to understand, honor, and integrate each others' perspectives:

> People in government often have ethical concerns with a randomized evaluation because you have one group who is getting an intervention, another group that's not getting it, and by randomizing it you are somewhat deciding who's going to which group... [We work to explain that] a randomized evaluation especially helps you understand the impact and you will be able to use that data and say 'my money is better spent on TaRL than on textbooks' and you're able to give causal evidence ... it actually helps you make better ethical decisions. - *TaRL researcher*

In contrast, interventions that did not explicitly identify and incorporate multiple stakeholder currencies into the design process inevitably broke down at some point of the intervention. Anticipated participation or support by key stakeholders fell through as professional, economic, or social constraints pulled them away, often despite their intentions. In the case of Aqua+, for example, only one type of scientific evidence and perspective—the most reliable way to use the solution to purify water—was identified as critical in the design of the product. Even at the product design stage, the perspective of peer organizations, the behavioral sciences, or potential downstream partners was not considered. Further, the needs, assets, and constraints of critical partners such as store owners or community leaders were not integrated, even when some actively sought to give feedback and engage with the company to improve the model—there was simply no mechanism or space for such perspectives to enter the process. For example, distributors had useful and reliable information on why the small bottle presentation was not working well for end customers or for store owners, but the relationship that TARA had established with them was arms-length, mediated by a distribution partner with no mechanism for information flow.

Mapping Learning Pitches. Negotiating a Shared Problem, Establishing a Learning Orientation, and Leveraging Distributed Expertise helped answer the questions of why and how actors would collaborate. A next critical step was to establish *what* exactly they would be doing. In successful cases, participants had identified that the intervention would face a variety of nested uncertainties and interdependencies.

And they had recognized that their traditional approach—and expertise—of planning for the linear implementation of a full design would likely fail.

When rock climbers face a large, complex climb, they break it down into a sequence of smaller pitches, typically scaled to the length of a rope, and punctuated by ledges on which they can stop, regroup, and redeploy their rope system for the next pitch. In similar fashion, once participants in our successful cases established a learning orientation and identified uncertainties, they jointly mapped a sequence of learning cycles—or learning pitches—to resolve them. Each cycle was mapped to resolve a relatively tractable, testable, and contained question. Like other programmatic scaffolding practices, Mapping Learning Pitches was actively deployed as a structured and facilitated practice, which systematically broke up what was otherwise an overwhelming range of uncertainty into discrete and interconnected questions that could be tackled sequentially.

Effective uses of this practice shared three characteristics: they were highly participatory, both to integrate distributed expertise and ensure that currencies of exchange were fairly traded; they were staged and iterative, isolating and sequencing specific uncertainties to address them in a systematic fashion, but also updating them as information from prior pitches emerged; and they alternated between a hyper-local and a global perspective in the evaluation of nested uncertainties. Each pitch was understood to bring together different strands of expertise and to be part of a larger chain of learning cycles. And the most experienced actors, like Pratham, learned to repeat the entire mapping process every time they entered a new context, identifying components that were more or less likely to "travel well" from one setting to the next.

Knowing that the next pitch, and the next one, would advance towards a coherent intervention provided a stabilizing frame that allowed actors to focus on a smaller, tractable pitch that required their full attention. Conversely, without this guardrail, actors in unsuccessful cases were simultaneously exposed to the entire constellation of uncertainties, leading to two related derailments. First, the range and amount of uncertainty could be overwhelming, leading to a sense that "everything depends on everything," such that conversations became repetitive and paralyzing. This, in turn, led certain actors to impose their perspective or for each actor to fixate on the component they knew best, imposing a siloed view of what should happen *there*.

Each stage of Pratham's development has been determined by what they have learned and what they are yet to understand at different levels of analysis, from individual children and teachers, to classrooms,

to schools, to districts. Even when an RCT had demonstrated significant gains in learning outcomes, Pratham insisted on adjusting and testing the model at each new site until the most impactful, consistent, and cost-effective components were identified. It also gave them clarity on which elements were easily replicated vs. which needed to be adapted from one context to the next:

> In our reading and arithmetic work we have put in very structured measurement systems, very uniform across geographies, across programs, which has really allowed us to understand where we are succeeding, where we have some problems, where we can make some modifications to improve impact without increasing program budgets. How do we reduce cost? There's this constant openness to experiment on different aspects... People often say, if you have something that is working, why don't you just keep expanding that? What happens at Pratham is once we know something is working, we look at the other pieces to see how we can make it better. - *TaRL implementer*

Naturally, learning pitches don't always go as planned and they produce unexpected results. Yet, as a senior Pratham manager explained, "We set very high standards for ourselves, but we are not scared to fail. That's a cultural thing in Pratham. If you do fail, it's okay, as long as you acknowledge it, understand what happened, and make a change from it. It's fine."

For example, between 2008 and 2010, Pratham and J-PAL evaluated two government-implemented partnerships that scaled Read India to the state level in Bihar and Uttarakhand. The RCTs tested Pratham's approach when implemented by government teachers (instead of Pratham staff), supported by training and materials. The results showed that the methodology failed in both states. Qualitative data, including interviews and process monitoring, identified that teachers had not fully complied with the model. They believed in its promise, but saw it as outside of and taking away from their core responsibility of completing the official syllabus. In response, Pratham and J-PAL then worked with the Haryana government on a new RCT that integrated TaRL into teacher responsibilities. They gave full ownership of the program to the government, and supported the training of "cluster resource coordinators" who monitored and mentored teachers. The government added an extra hour of TaRL instruction to the school day, grouping children by learning level. This new RCT revealed gains that exceeded expectations, driven by the effectiveness of cluster resource coordinators in driving teacher compliance. The next stage would be to learn how to effectively weave these practices into local government policy:

> To move from an academic paper to informing policy as we see it every day and making it useful, especially for governments. That was the space for the next role for J-PAL, because things don't happen by themselves and we have not learned to do that well. - *TaRL researcher*

In contrast, TARA did not identify uncertainties of mapped learning pitches. They framed the problem as a "product" question, focusing only on purification effectiveness and marketing. TARA did not recognize the importance of examining the needs, preferences, and habits of customers and independent store

owners and factoring them into the design of the product and its distribution channels. This was partly driven by blind spots created by TARA's prior successes and their choice of business model. Thus, . In hindsight, and employee said "...the technology was only 20, 25% of the puzzle. 75% of the effort should have been to understand how to connect with the customer, and we got it backwards."

Based on its original framing, TARA interpreted disappointing sales as a lack of customer and community understanding of the health benefits of Aqua+. There was no consideration of alternative definitions of the problem. For example, to improve sales, TARA asked its distribution partners to identify a network of micro-franchisees who would push the product directly to households. Simultaneously, TARA sought to expand beyond Delhi through partners in other states. They also launched an ambitious social marketing campaign to communicate the health risks of drinking untreated water and the benefits of Aqua+. The product design and distribution frequency, which were the actual key issues, were nevertheless maintained and sales remained well below expectations.

Scaffolding Practices: A Process Model of Evidence Integration

As detailed above, while they each played an important, independent function, the structural and programmatic scaffolding practices operated in a dynamic and interdependent fashion to provide the stability that participants needed to support a collaboration. The practices jointly and progressively allowed participants to reach the basic agreements that they lacked—by virtue of belonging to different organizations, sectors, and communities of practice—on the fundamental questions of organizing. These included which actors would be involved, what brought them together, where they were going, what approach they would follow, how they would distribute work, how they would coordinate effort, and how they would assess progress. In this section and through Figure 1, we propose an integrated model that describes the dynamics between the scaffolding practices, as well as a generalized sequence in which they often operated in our successful cases.[5] For reference, Table 4 summarizes the functions provided by each scaffolding practice.

A frequent, almost universal, point of entry was when an actor (most frequently a policymaker or practitioner) identified an opportunity to tackle a problem that they identified as difficult because of its scale, its resistance to prior attempts, insufficient or contradictory knowledge, the number of actors involved, and the fact that it was interconnected with other problems—or what is often labeled a 'wicked problem.' Opportunity windows could open because of political shifts, such as the beginning of a new government administration; a crisis that brought temporary attention and urgency to a longstanding issue; or the

opening of a novel and ambitious financing effort by a large funder. For example, TARA saw the open invitation from Antenna Group as an opportunity to bring its expertise and resources to tackle one of India's largest and most policy-resistant health issues.

In all of our cases, those who identified an opportunity recognized that the scale, complexity, and history of the problem called for unprecedented collaboration between stakeholders across communities of practice and sectors. The scale also prompted them to seek existing evidence, or what was known about the problem and prior attempts at solving it. From this common starting point, successful cases showed important similarities in their approaches and sequence. In contrast, and as shown in Tables 5 and 6, there was more variation among less successful cases, even though they faced similar challenges. From these common patterns we constructed the model that follows.

In our successful cases, those who identified the opportunity chose to first convene a diversity of stakeholders to jointly analyze the state of knowledge, rather than frame the problem themselves. This often began as a series of informal conversations to identify experts who could be knowledgeable, enthusiastic, and open partners. These conversations generally evolved gradually, adding experts to the mix that further clarified the breadth and complexity of the problem. As the conversations grew in depth and seriousness, actors identified a need to formally and jointly frame the problem to identify potential paths forward. Given their lack of a shared organization or frame, this need called for a formal convening space which, as mentioned before, could take different forms but often started as a small conference or seminar that was, by design, interdisciplinary and with the explicit goal to crystallize a joint definition of a problem and potential approaches to address it.

The challenges of agreeing on a joint frame given evident differences in ethos, approaches, and incentives of the different stakeholders created the immediate need for two scaffolding practices. On the programmatic side, participants needed a structured process and framework to ensure that their different perspectives would be integrated in a balanced and fair way, avoiding the imposition of a single agenda or perspective. The facilitated process of Negotiating a Shared Problem emerged as a solution. And, given the need for expert facilitation that could genuinely understand and mediate across perspectives, an actor naturally volunteered or was actively recruited to provide Multivocal Orchestration. Typically, this was an actor from one of the stakeholder groups with significant facilitation experience and an interdisciplinary professional trajectory.

Thus, in the first, problem framing stage of the design process, actors who attempted an inclusive, integra-

tive approach encountered the need to convene under a protected space, recruit multivocal orchestration, and establish a structured negotiation to identify joint objectives and define shared norms of engagement.

As actors moved from problem framing to framing of potential solutions, new challenges emerged. First, Negotiating a Shared Problem typically also established that the collaboration entailed multiple, significant uncertainties and would need to be equitable, without a hierarchy of interests but rather a dialogic approach to collaboration. Yet, it was clear that partners simply did not have enough information to agree on a solution. Instead, they followed a structured process to identify all that was (not) known regarding the different elements of the shared problem. Through this identification of the different types of uncertainty, participants were faced with all that they had yet to learn, which not only guided them to establish a learning orientation, but also revealed the need for additional scaffolding practices required to coordinate the complex work ahead. They could all agree that the learning collaboration would require them to extend beyond their traditional roles and organizational boundaries.

Given the challenges raised by a lack of established organizing agreements, participants could not rely on a centralized hierarchy, so they explicitly negotiated who should have more authority over which aspects of the problem, identifying expertise to distribute authority. But as this negotiation highlighted disagreements and fundamental differences in priorities and expertise, it revealed the need for a shared framework to settle disputes and disagreements. Through continued multivocal orchestration, this led to an explicit negotiation and definition of shared values and priorities that would guide the settlement of disputes. With clarity in shared values and distribution of authority, participants were able to settle on an agreed process.

As participants engaged in process design, they quickly began facing the constraints and trade-offs of a collaboration that necessarily extended them beyond the roles, organizations, and communities of practice that they belonged to. As actors experienced tensions, they anticipated that more would inevitably arise–for themselves and others. Their incentives, jobs, and demands were simply not aligned with the work required by the intervention. Concerned about their collective ability to sustain the collaboration, participants actively negotiated the currencies of exchange that would allow them to justify and continue their engagement: As participants expressed their needs, they also collectively and explicitly identified exactly what each brought to the table and what each would need to get in return (and in what time frame). Thus, while the need for equitable, dialogic coordination without a shared organization created challenges around the fundamental questions of organizing, the programmatic scaffolding practices affirmed a new

set of organic agreements and incentives that provided clarity and stability for participants.

Once participants reached organizing agreements on why they were working together, where they were going, who would be required, and how they would distribute authority, they faced the question of exactly what they would need to do and in what sequence. As they shifted from process design to the design of prototypes for different components of the problem (or the "actual work"), they remained oriented to the collective learning needed to navigate unresolved uncertainties. But, given the lack of centralized authority, a critical question was how to sequence activities and how to hold each other accountable for work performed in them. This question had three dimensions: a) the amount of uncertainty that could be tackled at a given moment by the collaboration—and the participants most responsible for each component—, b) the sequence in which different uncertainties should be addressed, and c) the type of evidence that they should seek at each stage to effectively inform subsequent choices. Jointly mapping learning pitches provided a sequence of tractable subprojects with clear guidelines for the distribution of authority and linked to specific learning objectives. As the model shows, each learning pitch in our successful cases was followed by a collective convening under a protected space to share and assess results, adjust course as necessary, and map out the next pitch.

Thus, as reflected in the model, the cycle repeated itself after each learning pitch. Some cycles only required a small revisiting of the programmatic scaffolding practices to reconfirm agreements and alignment. In other cycles, however, the learning pitch revealed the need for a formal redefinition of how the partners had understood the problem, their planned subsequent pitches, and the agreements that would be required to support them.

Ultimately, the agreements that were facilitated by the programmatic scaffolding practices directly provided the necessary stabilization required for actors to engage in the extensions demanded by the work. Arguably, these programmatic scaffolding practices could have been sufficient conditions for the integration of evidence into practice. But the likelihood, strength, and sustainability of their emergence and efficacy were dramatically amplified by the structural scaffolding practices. Indeed, we did not observe any cases where the programmatic scaffolding practices emerged without their structural complements.

This especially includes Adaptive Financing partners. In some cases, the same partner played a central role throughout the entire intervention, but in others financing partners had to be strategically brought in at different junctures in the process, to prevent them from dominating the agenda. Including an adaptive financing partner in convenings under protective spaces could provide valuable opportunities to affirm

new norms around purpose, knowledge, authority, and the importance of investing in learning. At the same time, if interested financing partners were unwilling or unable to let go of prescriptive, output-based orientations, an interventions was better off shielding the initial stages of the process from them.

In theory, the extensions and stabilizations provided by structural scaffolding practices could be deployed to support other types of interdisciplinary collaborations, which may required a different set of agreements and thus a different set of specific programmatic scaffolding practices. In our cases, the three structural scaffolding practices facilitated and amplified the emergence and efficacy of the programmatic practices by providing an overarching stabilizing framework.

Discussion

This paper studies the integration of evidence into policy and practice in the context of international development, which constitutes a particularly rich and consequential instance of knowledge transfer across *both* communities of practice and organizational boundaries. The ability to share, coordinate, and transform knowledge across organizational and occupational boundaries is increasingly relevant for organizations, because of the changing nature of work and the permeability of professional, organizational, and work boundaries. Most important, some of the most compelling problems of our time fall into the category of "wicked problems," which are large, complex, partly unknown, and in constant change. They present a heightened degree of complexity and uncertainty. And they sit at "intersections of intersections" between occupations, organizations, sectors, cultures, and national boundaries. Traditional approaches to problem solving are often fragmented across communities of practice and organizations, so they cannot suffice (Brynjolfsson and McAfee 2014; Leonardi and Bailey 2008; Benner and Tushman 2015). There is an additive effect as the uncertainty faced by each actor in their area of expertise is added to those faced by others. But there is also a multiplicative effect, as there is also uncertainty about how we should interact with each other and how our actions will affect—and be affected—by the evolving context.

Many barriers to knowledge integration in such complex settings have been well identified. And research has explored how coordination emerges in situations of occupational embeddedness, where actors faced multiple nested goals, distributed accountability and decision making, and high levels of uncertainty (Bechky and Chung 2018; DiBenigno 2018; Kellogg et al. 2006). A consensus has emerged on the need for a process of flexible and close-knit, "dialogic coordination" that is "highly situated in the specifics of the unfolding events (...) and occurs at the boundaries between communities of practice" (Faraj and

Xiao 2006).

As we document throughout the paper, the agreements, steps, and actions required for such dialogic coordination were indeed critical, but they were also risky, unusual, and complex. Successful actors in our study exhibited a combination of unexpected and potentially risky mindsets and behaviors that were determinant for success. ((These included (a) unusual degrees of curiosity and humility, rooted in acknowledgment that each actor only possessed a partial, limited, and contingent understanding of the problem; (b) extraordinary inclusiveness and a willingness to engage and work across organizational, occupational, and hierarchical divides; and (c) nimble adaptability, releasing actors' attachments to previously held beliefs about what worked, recognizing setbacks as moments of collective learning, and pivoting in their approach in response to emerging evidence. Effective dialogic coordination, therefore, demanded significant "leaps of faith" from actors, which required additional processes and structures of support (Pratt et al. 2019).

A critical challenge in our cases, however, was that participants belonged to different communities of practice and organizations, so they lacked baseline agreements around the basic questions of organizing that actors within the same organizations (and often communities of practice) can rely upon and take for granted. Thus, in order for actors to engage in the extensions and leaps of faith required for collaboration, they needed to construct an alternative set of agreements to provide the stabilizing effects of collective coherence and commitment, clear mechanisms of accountability, and a segmentation and sequencing of uncertainty into tractable stages.

Scaffolding Practices that Enable Collaborative Learning

Through contrasting international development interventions where integration of knowledge across boundaries did or did not happen, our study contributes to this line of research in three broad ways. First, our analysis uncovered the importance of a set of interlinked scaffolding practices—temporary, modular, and interdependent structuring practices that were intentionally deployed to help actors "solve a problem, carry out a task, or achieve a goal which would be beyond their unassisted efforts" (Valentine and Edmondson 2015). In our cases, each practice built on others in service of two critical functions. The first was an *extending* function, which led actors to take the needed "leaps of faith" and adopt mindsets, engage in behaviors, and establish ways of relating to others that were counter-normative and went beyond their professional or organizational mandates. Each scaffolding practice also played a counterbalancing *stabilizing* function that provided protection, stability, or compensation so actors felt safe, free,

and motivated to engage in the necessary extensions.

The analysis of the specific extending and stabilizing functions that each practice played at different junctures adds nuance and clarity to our understanding of certain practices that had already been identified in previous work, as well as the relationship between them. For example, previous work has documented how knowledge transfer and cooperation across communities of practice requires a shift in how participants understand their professional identity (Lifshitz-Assaf 2018; Kellogg 2011). Our paper builds on this work to identify the initial conditions that promote this identity shift, as well as the support needed to maintain it. Key among these was the periodic, facilitated Convening under a Protective Space, which operated as as a liminal area (Turner 1969), liminal space (Howard-Grenville et al. 2011), or relational space (Kellogg 2009) that invited the temporary suspension of habitual hierarchies and identities, creating the room for a new shared identity (Canales 2016; Kellogg 2009). Our work further clarifies the explicit work required to create such relational spaces, as well as the constant structuring activities required to monitor, maintain, reconfigure, and redeploy them as the nature of the collaboration evolves.

Our study also highlights that relational spaces, while critical, only provide the potential for identity suspension and reconfiguration. Particularly when participants lack the baseline understandings and agreements that can be taken for granted within a shared organization. To generate and sustain coherence and direction without a shared hierarchy or a predefined solution, actors resorted to Negotiating a Shared Problem, which crystallized the collective aspiration, identified a set of shared values, and established a unifying measure against which learning could be held. But participants could not get there on their own. In our successful cases, the diversity of perspectives and approaches presented by participants meant that negotiations required expert Multivocal Orchestration to facilitate interactions between them.

The importance of boundary work (Gieryn 1983; Zietsma and Lawrence 2010), brokers and boundary actors (Lingo and O'Mahony 2010; Bechky and Chung 2018), or boundary organizations (O'Mahony and Bechky 2008) has been well documented. We contribute to this line of work by identifying the variety of roles and extra work that multivocal orchestration requires—including overall project management, the deployment and facilitation of different programmatic scaffolding practices, and constant scanning of the needs and concerns of different stakeholders—as well as how its effectiveness is interdependent with other scaffolding practices. Notably, Establishing a Learning Based Orientation was a scaffolding practice that not only set the tone of the broader collaboration but also served as a constant frame to interpret emerging information and help orient future action.

Beyond detailing how known scaffolding practices operate and interact, this paper also documents new, critical scaffolding practices. Key among them is the structuration of Adaptive Financing, which can uniquely support (or, in its absence, hinder) knowledge sharing across communities of practice. As we document in the paper, such Adaptive Financing must strike a balance between flexibility to provide space for experimentation and change course as learning emerges, and accountability to observable, valuable *learning* outcomes. When not present as an actively structured practice, financing is typically rigid and attached to specific expected outputs, which can severely limit actors' ability to experiment and learn. Without accountability to learning, financing ban be *too* flexible, becoming disorienting and wasteful.

To guide learning in the midst of expansive uncertainty, actors in our successful cases engaged in an interative process of Mapping Learning Pitches. This scaffolding practice allowed participants to limit the amount of uncertainty they tackled at a given moment, identify the relationship and sequence between different types of (nested) uncertainties, and better focus on a narrower set of learning goals for each stage. This scaffolding practice, in turn, rested on a thorough process of Pricing Currencies of Exchange that clarified what each actor brought to the table for each learning pitch, as well as what they would need in return to remain meaningfully engaged. Reminiscent of *trading zones*, the identification of this practice helps clarify how such mechanisms of trading are explicitly deployed and facilitated, as well as the different dimensions—including organizational incentives, unique resources, non-negotiable needs, evidence priorities, time cycles, and target audiences—along which trades actually occur.

While Convening under Protective Spaces temporarily shielded actors from their habitual identities and pressures, Pricing Currencies of Exchange established strategic tethers to unavoidable outside realities that allowed each actor to sustain their participation in time. It also had the added benefit of establishing empathy, understanding, and trust among participants as each of them was pushed to ensure that the intervention met the needs of others. It thus worked as an effective structuring process to initiate and sustain "anchored personalization," or the types of trusting, long-term, personal relationships required to collaborate across communities of practice (DiBenigno 2018).

An Integrated, Inhabited Model of Knowledge Integration

A second, related contribution of our paper is that the analysis of the process through which each case developed and the critical decision points of each case allowed us, through contrasting counterfactuals, to develop an integrated model that describes how all the different scaffolding practices build and depend on one another. this model clarifies the role that each scaffold plays in relation to the others and why,

in combination, they are so effective. It details how structural scaffolding practices create the space and preconditions where programmatic scaffolding practices can be deployed. It also suggests a general sequence in which they often functioned. Most important, the model highlights the extent to which complex, intersectoral learning collaborations lack the types of baseline agreements that can be taken for granted within shared organizations. They thus require the active and strategic management of scaffolding practices that guide actors to the establish shared agreements and understandings necessary for dialogic coordination. Our paper unpacks the specific things that actors do, at specific junctures, to provide temporary support (through scaffolding practices) and bring others along in the complex tasks of learning across boundaries.

Ultimately, the goal of each of our international development interventions was to create new social structures—institutions, programs, and organizations—to better address the needs of those most in need. Like any building effort, the work to establish new social structures requires the joint, collaborative effort of many individuals working outside of what has existed before. But given the amount of uncertainty and absent a shared formal organization, actors also require temporary supporting structures—scaffolding practices—to guide and contain construction efforts.

Far from being inert, scaffolding practices had to be deployed, managed, and constantly reconfigured to provide enough stabilizing factors to allow participants to engage in the extensions demanded of them. This again highlights the importance of Multivocal Orchestration that constantly straddled between participation in the actual activities of evidence integration and the structuring work required to identify, deploy, and adjust the different scaffolding practices. In that sense, multivocal orchestration bent beyond facilitation and management and required a fair amount of sociological imagination (Mills 1959) or sociological citizenship (Silbey 2011), to identify and tend to the structural needs, constraints, and dynamics of all participants. It was through careful multivocal orchestration that collaborations identified the lack of a shared set of basic agreements on the fundamental questions of organizing, which required the establishment of a new set of shared agreements to enable collaboration.

Knowledge Integration across Communities of Practice as Navigation within Uncertainty

As a third contribution, our analysis clarifies the ways in which *uncertainty* is the central driving force in the types of complex, intersectoral collaborations required for evidence integration. Indeed, the common

catalyst that led certain actors to identify and deploy scaffolding practices was the recognition of the amount of complexity and uncertainty surrounding the context they were entering. As described above, these were contexts with systemic complexity, subject to constant change, involving multiple stakeholders, and beyond the scope and knowledge of any single actor. And where many prior, capable efforts had failed before. the humble assessment of the terrain and all that was now known about it convinced actors that an entirely different approach to problem solving would be required. It was from this recognition of uncertainty that some actors started a new effort with a convening of a diversity of perspectives focused on learning. The focus on joint learning demanded new ways of working together, which came with its own set of relational and procedural uncertainties. It was the recognition of this second layer of uncertainties, amplified by the lack of a shared organization and its accompanying set of agreements, that led certain actors down the path of enabling structural supports, seeking to provide enough stabilizing agreements to enable the types of extensions needed from participants.

In contrast, every one of our failed cases can be traced back to the underestimation, oversight, or neglect of one or several sources of uncertainty. The resulting approaches were insufficiently inclusive, experimental, incremental, flexible, or adaptive. To understand why this is the case, it is illustrative to further unpack the dynamics of uncertainty within the cases that best identified it.

As described above, when actors entered a process with awareness of uncertainty, they identified that there would be different types and sources of uncertainty, which combined in additive and multiplicative ways. As they sought to categorize the different unknowns and the sequence in which they could be tackled—through Mapping Learning Pitches—they planned for a somewhat linear learning process where pitch A would inform pitch B and so on. Even in the most thoughtful and inclusive of processes, however, actors soon discovered that their initial assessment had underestimated the actual nature and amount of uncertainty. As a result of each set of experiments, therefore, actors often expanded their original assessment of the different remaining uncertainties. They also inevitably diverted from the linear path they had traced, often in unpredictable ways. This initial learning further cemented the need for gradual, iterative approaches that would tackle uncertainty in a systematic and ordered way, which further reinforced the need for scaffolding practices to enable adaptability and learning throughout the meandering process.

Conclusion

Despite the widely accepted imperative to adopt an evidence-based approach, development practitioners often bring entrenched mindsets and behaviors that constrain an adaptive and inclusive approach to learning and implementation. Faced with an array of organizational, political, financial, and logistical pressures, a top-down, output-oriented approach to problem-solving can offer the illusion of efficiency, measurability, and control, particularly in the absence of a clear alternative. Our work offers strategies that allow development actors to balance their need for accountability, efficacy, and predictability with flexibility, inclusiveness, and learning.

Our research explores the challenge of evidence-based practice within international development as an example of learning across occupational and organizational boundaries. This is critical because, despite the scale and urgency of international development issues and the recognized imperative of applying evidence-based practices, we currently lack an good frameworks to sustain them, particularly under conditions of complexity and uncertainty. In our research, successful evidence-based practice was enabled by a set of key structural and programmatic scaffolding practices that first provided enough stability around professional and personal norms—around questions of identity, purpose, and knowledge—so that actors could engage in unusual and risky behaviors required for inclusive and relational learning.

On a fundamental level, scaffolding practices provided actors with a set of guardrails that both supported and guided them towards intentional engagement with the complexity and uncertainty inherent in the international development context. Each of these practices interacted to help actors re-orient towards learning and dismantle rigid hierarchies of expertise in order to leverage diversity and collaboration. From this new approach to knowledge, actors were able to manage the complexity of international development through the proactive negotiation of incentives and timelines, as well as adaptive, partipatory responses to evidence as it emerges on the ground.

Notes

[1]By development practice, we mean the work of government actors, NGOs, and others who are responsible for designing and executing development projects and programs.

[2]Saturation is often difficult to define and is dependent on the field of study. In this case, we defined it as the moment when, in a sequence of several expert interviews, no interviewee gave us information that we had not encountered before.

[3]We ensured consistency in coding across the different cases and authors through several mechanisms, including: a) a selection of interviews was coded by two or more coders, after which they reviewed discrepancies and agreed on their resolution, b) a common project book where all the codes were collectively kept, aggregated, and analyzed, c) a weekly meeting to review coding process and to develop a joint coding standard, d) memos were developed jointly, with contribution from and verification by the different team members, among others. See (e.g. Evans et al. 2016).

[4]The two types of scaffolding practices still play important and distinct functions, as can be seen in cases where one was present but not the other (see Table 5).

[5]In this model, scaffolding practices operate as *subprocesses*, similar to Hannan and Freeman (1989, pp.60, 148-150), or separate elements of a process that build on and require each other to function. They mostly operate sequentially, but sometimes occur in parallel or in a different order.

Table 1: Expert Interviews

	Researchers	Funders	Intermediaries	Policymakers	Implementers	Total
Phase I: February 2015 - May 2016	10	8	6	2	5	31
Phase II: September 2016 - June 2017	10	7	6	3	8	34

Table 2: Case studies, stakeholders, and interviews

Name	Program Description	Dates of Intervention	Primary Stakeholders	Interviews
Teacher Community Assistant Initiative (TCAI) (Ghana)	Remedial education program in reading and math for primary school children, grouped by learning level instead of age, through after school programs delivered by teaching assistants from local communities or public school teachers. Main goal: scale in Ghana	2010 - 2013	Researchers, government officials at different levels	30
Graduating Ultra Poor (GUP) (Ghana)	Poverty alleviation program integrating elements of social protection, cash transfers, livelihood development, and financial services. Part of a network of RCTs conducted in parallel in several countries. Goal: Rigorous test of graduation approach in several countries.	2010 - 2013	Researchers, intermediary, NGOs.	
Teaching at the Right Level (TaRL) (India)	Remedial education program in reading and math for primary school children, grouped by learning level instead of age, generally taught after school. Goal: Design program, scale into national policy.	2001 - present	NGO, researchers, government officials at different levels.	51
Aqua+ (India)	Water purification drops for retail sale to base of the pyramid customers. One of several country-level adaptations of a particular water purification technology promoted by Antenna, a Swiss-based global NGO.	2010 - present	National, international NGOs, distribution network.	
Progresa - Oportunidades (Mexico)	Poverty alleviation program using conditional cash transfers, provided to heads of households conditioned on continued enrollment of children in school and demonstration of appropriate child healthcare and nutrition. Goal: Design program, scale nationally, change how programs are evaluated.	1997 - 2018	Government officials at different levels, academics, international agencies.	38
Programa Primer Empleo (PPE) (Mexico)	Government subsidy to promote creation of new jobs in the private sector for first-time job seekers. Subsidy granted to social security fees paid by companies for every employee hired. Goal: Scale in Mexico.	2007 - 2012	Government officials at different levels, business sector, Social Security System.	
CALIE (South Africa)	Employment program introducing new elements to vocational training, based on insights from behavioral theory. Included toolkits for vocational trainers, checklists for employee seekers, and recommendation networks. Goal: Improve existing policy, scale nationally.	2011 - 2016	Government officials at different levels, researchers, intermediary.	42
FUEL (South Africa)	School nutrition program. Sought to address issue that funding for school lunches was not translating to every child having a meal at school on time. Initial premise was that challenge was lack of logistical sophistication by education ministry. Later pivoted to focus on organizational capacity at individual school level. Goal: "fix" inefficiencies in system.	2007 - present	NGO, government officials at different levels.	

Table 3: Summary of Outcomes and Markers of Success

Name	*Evidence as evolving & unsettled (vs. settled in hierarchy)*	*Inclusive, Dialogic Coordination (vs. hierarchical)*	*Accountability to staged learning (vs. to implementation)*	*Project management as iterative adaptation (vs. linear implementation)*	***Summary:** Evidence Integration Result*	*Evidence Score*
Teaching at the Right Level (TaRL)	Yes. As part of their routine process, Pratham undergoes a constant update of what is known vs. remaining questions. Learning has now shifted from demonstration of methodology for remedial education to resolving uncertainties around avenues and levers for executing at scale in different environments. Throughout this process, they have a very clear identification of mechanisms that *have* to be present for program to work, vs. elements that *can* be adapted based on local characteristics, vs. elements that *must* be adapted to local needs.	Yes. From the origins of Pratham, knowledge is jointly generated with local actors from each context, through two approaches: a)Pratham is asked to bring their framework to new setting; b)they look for places with potential and "push" the approach, carefully adapting methodology through gradual, iterative process. Regardless of the type, they create a management committee with key stakeholder representation for each project. They also integrate findings from each endeavor to their body of knowledge. Follows full cycle each time.	Yes. TaRL has identified key questions and types of evidence necessary at different levels of analysis, from individual learning mechanisms to administration of intervention to scale up (curriculum vs. delivery of content vs. management practices vs. scale up at national levels vs. national evaluation of students). Each level requires partnerships with different institutional actors, each with own timeline, including academic institutions, international agencies, local, regional, and national governments, and donors.	Yes. Commitment to learning entails processes that are iterative by design and that evolve in cycles. Things are tested at each level of analysis through several iterations before scaling and integrating to next level of analysis. When new evidence emerged that challenged core assumption of the approach, for example, all Indian programs were rolled back until contradictions were resolved.	High. TaRL well suited to **scale**, systematically influenced education system **(spread)**, internalized evidence generation and use **(structure)**, improved learning outcomes for millions of children in India, and influences programs beyond India.	Sc: 2 Sp: 2 St: 2 - - **T: 6**
Progresa	Yes. From extensive review of existing research identified what had NOT worked and key uncertainties (e.g. whether conditionality is necessary, what type of conditionality, if transfer should be in cash or in kind, which agency could deliver last mile...) Design process started *from* those uncertainties, followed by the iterative development of the program, designed to resolve key uncertainties in sequence. Split design and implementation phase so different government agencies resolved different components (surveys, pilots, experiments). For example, for initial pilots partnered with local governments eager to learn. Ongoing evidence generation and learning built into design & implementation.	Yes. Carefully mapped entire theory of change and identified all necessary actors at each stage—including different levels of government (state, municipal, federal), academia, NGOs, international organizations. Steering committee with representatives from each group established to make all decisions. Two former academics recruited as government ministers to ensure learning and dialogue across sectors. Academics and international agencies recruited from the beginning as learning partners.	Yes. At every stage accounted for and managed complexity of stakeholders and political/social/economic variables. Started program with discretionary government funding to avoid rigidity of national budget cycle. Held off on scale-up to avoid interference with midterm elections. First partnered with friendly local governments to experiment with budgetary and implementation dynamics. Then designed national roll-out. Secured international loan to formalize lessons into scaled up program and remove temptation for political use.	Yes. Adaptive at every stage. Program was first housed at President's office to use only discretionary funding. Engaged taskforce from different ministries: different teams worked on parallel processes to learn which ministry was best suited for a national program. Ran several pilots with local governments to resolve key uncertainties in program design and implementation. Invited academic partners to evaluate every stage and adapt design before scaling further. Created a government evaluation function.	High. Progresa's evidence-driven approach succeeded on multiple levels: Program reached massive **scale** in Mexico; the concept of conditional cash transfer programs **spread** to dozens of countries; and program established CONEVAL, which now assesses the effectiveness of *all* government social programs through rigorous evaluation **structure**. Program transformed how Mexico integrates evidence into public policy.	Sc: 2 Sp: 2 St: 2 - - **T: 6**
CALIE	Yes. Program started as joint initiative to learn effective ways to reduce youth unemployment. Undertook joint identification of what was (not) known, and the key uncertainties that would need to be resolved before scale up (e.g. was bottleneck in candidates not seeing right jobs, or a mismatch between skills and jobs, or candidates not applying effectively to jobs?...)	Yes. Initial open consultation process to identify interested and necessary partners across sectors. Started as an open collaboration between government agencies, local and international academics, international NGOs, funders, and local implementers.	Yes. Explicit incorporation of political, labor market, and academic cycles in program design. Constrained intervention to first experiment with fit within current government budgets and organization; then tested potential worker incentives. Strategically identified government as project leader to ensure political alignment and protect learning approach.	Yes. Process designed to iteratively integrate different types of lessons from different critical stakeholders. Process designed to follow three cycles: 1)small pilots, 2)larger experiments, 3)scale up.	High. At end of study period, J-PAL Africa, the PIs, and the DoL were discussing an expanded roll-out of related policies and practices based on project results **(scale)**. Government also developed new practices for evidence-based policymaking **(structure)**. Through regional convenings, CALIE has informed work in the continent **(spread)**	Sc: 2 Sp: 1 St: 2 - - **T: 5**
Teacher Community Assistant Initiative (TCAI)	Yes. Key uncertainties were identified from the beginning with respect to the adaptability of the education intervention, adjustments and operations required to implement at scale, and integration with incentives, workloads, and work processes of public teachers required for intervention to fit within educational system (e.g. how to assess learning level in these schools, whether and how teachers could fit an intervention into existing, mandated structure, ...).	Yes. Intervention sought to adapt proven concept at the scale required for a national educational system. This required different types of knowledge at different stages—researchers with knowledge of methodology, policymakers needed for scale, NGO to help with logistics of the intervention, teachers to deliver the program, funders willing to finance scale up. Each perspective was actively incorporated through steering committee that included skeptics, adversaries, and champions.	No. Program success was dependent on a particular political timeline and set of political factors, and did not consider uncertainties in implementation timeline. This led to the temporary hiatus of work from main government partner. Lack of mechanisms of continuity or of resilience meant that when politics changed it was impossible to restart the work (given turnover of key individuals and change in institutional priorities).	Mixed. Design of program was framed around learning from early implementation stages, but was not resilient or adaptive enough to withstand political shifts. When political timeline shifted and one of the partners was sidelined politically, entire scale up was killed. Financing was not secured to ensure or motivate continuity of scale up.	Moderate-Strong. Success in terms of impact on students, influence on government in its awareness of importance on remedial education and evidence-based approach, and influence on neighboring countries and development partners **(Spread)**. However, did not scale up directly due to collapse of process and lack of new, dedicated funding to support adoption **(no scale or structure)**.	Sc: 1 Sp: 2 St: 1 - - **T: 4**

Table 3: (continued)

Name	*Evidence as evolving & unsettled (vs. settled in hierarchy*	*Inclusive, Dialogic Coordination (vs. hierarchical)*	*Accountability to staged learning (vs. to implementation)*	*Project management as iterative adaptation (vs. linear implementation)*	***Summary:** Evidence Integration Result*	*Evidence Score*
Graduating Ultra Poor (GUP)	Mixed. On the design and implementation of RCT, knowledge was treated as uncertain and in need of testing and validation. On the incorporation of government as a useful partner, uncertainty was not adequately accounted for. The question of scaling was not incorporated as part of the intervention.	No. A shared understanding of the problem and collective goal was not negotiated. It was only framed as faithful implementation of one of several RCTs conducted around the globe to expand evidence of the graduation approach. Hierarchy led by IPA. Desire of local partners to plan for scale up not accounted for.	No. RCT was defined as the only relevant timeline, which dictated implementation plan. Partners were identified, but competing interests and constraints not properly incorporated. Relationships were arms-length, through formal contracts.	Mixed. RCT was appropriately designed for flexibility and learning and adjusted at several points. Overall objective, however, was only defined as RCT, with no flexibility for amplified scope.	Moderate. Contributed to global body of evidence on the Graduation approach, informing policies and practices for dozens of NGOs and governments **(spread, structure)**. But did not fully engage policymakers or implementers as co-creators, disabling ability to scale up in Ghana **(no scale)**.	Sc: 0 Sp: 2 St: 2 - - **T: 4**
FUEL	No (but later yes). Initial hypothesis of process centralization and optimization (not consulted with other stakeholders) led to a specific approach seeking centralized efficiencies of meal resourcing and distribution for schools. Only after this failed, implementers were faced with evidence that challenged approach and made uncertainties evident.	No (but later yes). Project had strong initial framing on operations: it was assumed that nutritional shortcomings in schools were driven entirely by inefficient logistics. Clear hierarchy, later shifted to committee. Schools and school system were first seen as bottlenecks, not partners.	No. Initial pushback from key stakeholders was seen as evidence that they were the bottlenecks. Implementation process ran into reality of political and academic cycles that had not been considered and was forced to change approach.	Initially, the design was not adaptive. But the founders were committed to finding a solution (vs. to an approach). This allowed for a second iteration with a radically different approach. However, this was only after significant effort, time, and money. A less committed funder would have killed the intervention.	High (but costly). Through flexible, committed funding, FUEL was able to weather its setbacks. The school nutrition program ultimately reached significant national **scale**, and FUEL has expanded its collaboration with the South African government to other challenges in the education system **(structure)**. But learning process was slower and costlier than necessary because of initial orientation. No **spread** of its lessons .	Sc: 1 Sp: 0 St: 2 - - **T: 3**
Aqua+	No. Good identification of technological and operational uncertainties, but did not properly account for retail and consumption uncertainties (e.g. shelf life, product fit with consumer habits, behavioral changes required for families to use drops at home, etc.).	No. Focus was on effectiveness of technology (droplets for water purification) and operations (efficiency in value chain), so technical knowledge deemed sufficient across contexts. Neglected last mile retail partners and communities as critical knowledge-holders to develop a sustainable business model and uncover insights about *use* of technology.	No. Program failed to incorporate complexity and diversity of timelines and stakeholder needs. Did not align learning cycles with Antenna and other partners. Did not consider timelines or needs of downstream actors (e.g. stocking dynamics for small retailers). Ignored alternative distribution approaches tested by Antenna partners in other locations.	Mixed. Program was adaptive in their core technology and upstream operations, but lack of focus on retail partners and consumers, plus financing structure dependent on early success created rigidities in key downstream components.	Low. Failure to give sufficient attention to incorporating evidence and recognizing different types of evidence **(little structure)**. Major adjustments in implementation undertaken only much later once issues were recognized **(no scale, no spread)**.	Sc: 1 Sp: 0 St: 1 - - **T: 2**
Programa Primer Empleo (PPE)	No. Took academic's theory of labor market rigidities as only piece of relevant information. Did not identify potential sources of uncertainty on assumptions or implementation. Did not create room for adjustment to potential lessons from implementation.	No. Intervention was designed by federal government without integration of other perspectives, including academics or civil society. Need for other levels of government was underestimated and business sector was explicitly ignored for fear of capture.	No. Incoming administration was focused on federal government's political timeline. Did not consider local governments or other important timelines, including budgetary cycles, midterm elections, or business cycles.	No. Ambition to scale quickly led to design for national roll-out, so program was inscribed in the federal budget from the first year. Strict government budgetary rules that mandate faithful compliance with stated use of funds removed ability to adjust program despite clear evidence for needed changes.	Low. Failure from policy standpoint **(no scale or spread)** but served as experimental exercise to understand complexities of policy design **(little structure)**.	Sc: 0 Sp: 0 St: 1 - - **T: 1**

Table 4: Extending and Stabilizing Functions by Scaffold

Type	Scaffolding Practice	Extending Function	Stabilizing Function	Related Literature (contributions in bold)
Structural	Convening Under Protective Space	Enables the co-creation of new identities, goals, **roles**, and norms conducive to evidence-based practice.	**Guards against destabilizing influence from home organization, justifies involvement,** provides continuity.	(Kellogg et al. 2006; Howard-Grenville et al. 2011)
	Multivocal Orchestrating	Leverages and synthesizes diverse knowledge and interests. **Guides extraordinary behaviors through facilitation**.	Facilitates fair negotiations, in service of both collective and individual interests. **Provides or secures effective project management of intervention. Manages, structures, deploys scaffolds.**	(O'Mahony and Bechky 2008; Obstfeld 2005; Lingo and O'Mahony 2010)
	Adaptive Financing	**Allows actors to engage in learning-based risk-taking and experimentation, without fear of losing resources.**	**Ensures resources will be provided at different stages of learning and implementation. Defines tangible measures of learning to ensure accountability.**	-
Programmatic	Negotiating a Shared Problem	Releases attachment to prior problem definition, integrates multiple perspectives, releases need for control.	**Generates collective coherence. Clarifies a shared goal, provides a shared language, and highlights a set of shared values.**	(Carlile 2004; Bechky 2003a; Valentine 2018)
	Establishing a Learning-Based Orientation	Reframes identity and expertise from implementation/output and promotes interdisciplinary discovery.	**Provides shared direction and metrics. Reframes failures as part of the learning process**.	(Lifshitz-Assaf 2018; Valentine 2018; DiBenigno 2018)
	Leveraging Distributed Expertise + Bodies of Evidence	Motivates risk taking and collaboration; establishes "heedful interrelating;" flattens "hierarchies of evidence" that differ across communities of practice.	**Establishes parameters/ authority of dialogic coordination. Affirms localized expertise by distributing relative authority.**	(Brown and Duguid 1998; Faraj and Xiao 2006; Weick and Roberts 1993; Valentine and Edmondson 2015; DiBenigno 2018)
	Pricing Currencies of Exchange	Recognizes, elucidates, and negotiates respective resources and constraints.	**Instills trust and buy-in among participants, justifies continued participation, as individual needs and incentives are identified and met.**	(Orlikowski 2002; Kellogg 2011; DiBenigno 2018)
	Mapping Learning Pitches	**Makes actors more willing to experiment with new behaviors and forms of interaction by delimiting uncertainty.**	**Allows actors to focus on a smaller, tractable pitch that demands their full attention**.	(Faraj and Xiao 2006; Valentine and Edmondson 2015)

Table 5: Presence of Structural Scaffolds by Case

Name	*Convening Under Protective Space*	*Multivocal Orchestrating*	*Adaptive Financing*
Teaching at the Right Level (TaRL) *Score: 6*	Yes. On multiple levels. For each question/challenge/setting, a steering committee was formed with all relevant stakeholders, including skeptics, critics, key participants. Pratham has established an independent organization to gather and disseminate evidence. Organizes conferences, workgroups, and task forces.	Yes. Pratham founders have straddled academia, public service, and business throughout their careers. Profoundly sensitive to trade-offs between policy design, policy implementation, and rigorous research. Pratham tends to be heavily involved in project design, management, and implementation of all its interventions (often remains behind the scenes).	Yes. On multiple levels. For each intervention, funds are raised with the explicit goal of learning, with flexible timelines, definitions of goals, etc. In addition, Pratham and J-PAL have raised additional, flexible funding to promote research, experimentation, and dissemination activities.
Progresa *Score: 6*	Yes. Multiple, evolving with the program. First, president's office provided political protection for initial design and experimentation. Later, a steering committee was created with participants from federal and local governments, academia, NGOs, and international organizations. It met regularly to evaluate progress and adjust program design. Finally, the program was institutionalized with a multi-agency steering committee and an autonomous government entity (CONEVAL) was established to rigorously evaluate the program.	Yes. President himself spent time in academia, as policy designer, and in government. He recruited two ministers from academia to be in charge of this program, as the intention was to move away from what had been proven to be ineffective and to create new policy paradigm.	Yes. Financing initially provided by discretionary budget from the president's office. Gradual budget increases were made until, after midterm elections, the program was finally established as a national program with federal budget allocation. International financing was also raised strategically and purposefully to ensure program continuity and integrity in the face of potential political change.
CALIE *Score:5*	Yes. Multiple, evolving with the program. J-Pal provided continuity, leading research. 1) Inter-sectoral conferences explored key strategic issues related to South African unemployment. 2) RFP launched, explicitly demanding multi-sector collaboration on evidence-based interventions. 3) Government-led research project was established with an inter-sectoral steering committee that met regularly.	Yes. J-PAL Africa acted as a convener and project manager, bringing together representatives from key stakeholder groups and establishing multi-sector steering committee. J-PAL's policy team were recruited explicitly because of multi-sector experience. While J-PAL managed process and project, government was strategically named project leader, especially for public communications.	Yes. Funding was raised for each stage/iteration with learning as the central objective. Program thus explicitly adapted to lessons from field. Funding rules included, for example, that program had to adapt to evidence but also to existing structure and funding of labor ministry to ensure adaptability to local conditions.
Teacher Community Assistant Initiative (TCAI) *Score: 4*	Yes. Conferences + steering committee was purposefully designed to include critics, champions, and skeptics. IPA took the lead as project manager. Factors for convening: a)TaRL's international recognition, b)Ghana as strategic site to test intervention at scale, c)international financing from high status agency, d) personal relationship between researchers and education ministry.	Yes. IPA team and government officials with academic backgrounds. IPA designed the interventions with the PIs, but coordinated implementation with education ministry through the steering committee. IPA served as mediator when evaluation and implementation were in tension (e.g. rigor of experiment vs. logistical feasibility for teachers in after school programs).	Mixed. Interdisciplinary team raised funding explicitly for learning, flexible to changing nature of project as lessons emerged. But financing for scale up was tied to a single government agency that had its own political cycles. When political timing changed, financing for scale up dried up and intervention stopped at the pre-scale stage (but with valuable learning).
Graduating Ultra Poor (GUP) *Score: 4*	Yes, but incomplete. Conferences + Steering Committee. 1) GUP was one of several country interventions to test the graduation approach. Belonged to working group that met regularly. 2) In Ghana, a convening space was created through IPA. However, it did not convene all stakeholders, particularly from government and NGO complementor, to consider scale up post RCT.	Mixed. IPA took unusual role, involved in both research design and implementation. Implementing NGO was hired as subcontractor (not as partner), but was given voice on key implementation issues (not on questions of sustainability or scale-up). IPA had tight control of all project decisions, coordinating stakeholders for RCT, but had limited engagement with central and local governments as partners.	Mixed. Funding for RCT was raised with flexibility in mind, to learn from early pilots and adapt implementation. E.g., training was changed after initial implementation. At the same time, opportunities to gain political buy-in and raise funding for scale-up was missed.
FUEL *Score: 3*	No [but later yes]. Intervention followed a top-down approach based on logistics expertise of founder. Because of status and recognition of founder, approval for implementation was granted by ministers with little input from lower levels.	No. Founder had background on food industry and logistics and framed the problem accordingly. Hired a logistics team and rolled out the initial intervention as a logistics program.	Yes, but not focused on learning. Funding was flexibile because of committed founder who persevered through several setbacks. The team learned from each iteration, but project structure was not set up for learning (i.e. they only adjusted at each iteration because it had clearly failed).
Aqua+ *Score: 2*	Mixed. The organization is part of an international network of country-level partners for Antenna. They convene periodically to share best practices. However, locally, the organization is split between basic research, experimentation, and implementation. Also, there was no inclusion of external stakeholders as partners.	Mixed. The organization is multivocal by design, integrating basic research, intervention design, and incubator to spin off independent organizations for each solution. But the core teams are part of separate organizations and, in the case of Aqua+, failed to integrate downstream considerations. Key external stakeholders were not included. No integration between areas/units, only hand-offs.	Mixed. Antenna provided flexible terms, with learning as the central aim of the multi-country approach. It systematically structures projects with flexible milestones to allow for adaptation to ongoing learning. But financing was not tied to *demonstrated* learning and lacked counterbalancing rigor. Strategy was to raise scale-up funding based on successful implementation of the approach. When it failed, subsequent funding was not realized.
Programa Primer Empleo (PPE) *Score: 1*	No. In fact, program designers sought to *shield* the program from outside influence under premise that outside stakeholders would capture and distort it. Program design carried out entirely by President's transition team.	No. Program was designed exclusively by government, who was skeptical of outside influence and opinions. Assumptions were taken from a single academic paper, but academics themselves were not engaged.	No. There were strict budgetary and operating rules that provided no flexibility. Due to pressure to scale the program quickly during the first year of the presidency to meet campaign promises, program was included in federal budget, as a national program, from day one, with rigid requirements.

Table 6: Presence of Programmatic Scaffolds by Case

Name	*Negotiating a Shared Problem*	*Establishing a Learning-based Orientation*	*Leveraging Distributed Expertise*	*Pricing Currencies of Exchange*	*Mapping Learning Pitches*
Teaching at the Right Level (TaRL) *Score: 6*	Yes. Each intervention or question started from a local coordinating partner who built a steering committee or task force and negotiated a problem statement with all relevant stakeholders. Often, negotiation framed around new grant proposals.	Yes. Pratham's founding principle is to learn how to improve education outcomes, wherever the best evidence may lead. It has strong and longstanding partnerships with academic and research institutions, including J-PAL. Each stage of the firm has been defined around what has been learned and what is yet to be understood. Has created specific organizational units to design, implement, and disseminate research.	Yes. Pratham creates steering committees for each intervention, where stakeholder closer to each element of the methodology is given disproportionate voice. Unusually open to mix of research methods, approaches, and partners including RCTs, theoretical papers, case studies, large-scale data, community surveys. Pratham has a specialized unit that integrates different lessons and disseminates to different audiences.	Yes. Pratham has learned to balance tradeoffs between political needs and cycles (e.g. giving disproportionate credit to government partners), implementation realities (needs, rhythms, and constraints of teachers and schools), rigorous research (counterfactuals and control groups, different types of evidence for different questions), and donors (judicious use of resources and evidence of outputs).	Yes. Commitment to learning and rigorous experimentation, systematically testing key variables to determine essential levers of impact, and elements to adapt for different contexts. This allowed them to be flexible to meet stakeholders' needs, critical to secure long-term buy in from policymakers and other partners. Co-design with local partners each time to gain buy-in and leverage distributed expertise.
Progresa *Score: 6*	Yes. Began with broad consultation with academics, international agencies, and officials from different countries to identify what had *not* worked. During pre-design, identified cash transfers as a promising avenue, but one that needed appropriate conditions and methods of delivery. The shared problem was defined as "eradicating *trans-generational* poverty," focusing on child health and education.	Yes. The program was equally oriented towards generating evidence and insights, as on achieving outcomes. The program's commitment to learning was reinforced by the explicit assignment of resources to reviewing and internalizing evidence, as well as dense web of communication that took place among stakeholders on a weekly basis.	Yes. Systematically engaged academics, international agencies, government officials, local governments, NGOs. Steering committee designed as collegiate body where key stakeholders had voice and veto power. Multi-agency and multi- sector steering committee ensured most knowledgeable member had disproportionate voice for each component of the program.	Yes. Needs and assets of political and academic stakeholders were identified and met. Academics were given unrestricted access to field and data to ensure a rigorous evaluation. Launch of national program was postponed to avoid interference with midterm elections. International agencies were engaged to provide institutional backing and program continuity in face of political transition.	Yes. Evidence demonstrated that *no* generalized subsidy had ever worked. Some mixed, inconclusive evidence supported targeted and focused approaches. The development of the program was designed to systematically and sequentially uncover key uncertainties and test critical assumptions through a three year design processs that would culminate in the development of a national policy.
CALIE *Score: 5*	Yes. First through broad, multisector conferences where key strategic issues for youth unemployment were identified. Negotiation framed around grant proposal. Then, task force was established that explicitly incorporated skeptics and critics as well as champions to define a shared, compelling problem around youth unemployment.	Yes. All stakeholders were equally committed to the problem and to the value of learning, even if their desired outcomes were not realized. Funding was not dependent on success of the project but, rather, rigorous learning and transparency.	Yes. Steering committee included every stakeholder, including the labor ministry, labor centers, academia, and local NGOs. J-PAL's project management ensured coordination. For each element of the program the most experienced stakeholder was given additional weight.	Yes. Each stakeholder began the process by describing their constraints, desired outcomes, available resources, etc. Each became a binding constraint for the design process. For example, the intervention had to be implemented within the existing structure and budget of the government's labor centers. Strategically, government was identified as leader of project to gain political support.	Yes. Identified key assumptions and uncertainties as well as key constraints. Designed intervention as a sequence of iterations that built on each other to address each uncertainty through best suited methods and at the most appropriate scale (small experiments and pilots to test key programmatic assumptions, larger experiments to test both program design and ability to implement at scale).
Teacher Community Assistant Initiative (TCAI) *Score: 4*	Yes. Steering committee included key stakeholders, including skeptics and challengers, that were included in the definition and understanding of the problem. Problem negotiation organized around grant proposal. Project proposal was designed to seek alternatives to remedial education based on TaRL evidence.	Mixed. Intervention was designed to tackle key uncertainties in sequential cycles. At each stage, lessons from the field were incorporated. But learning-based approach did not include political cycles or changes, so scale up did not occur.	Yes. Created multidisciplinary steering committee with all stakeholders. Each stakeholder was given disproportionate voice for key component (e.g. academics for evaluation, teachers for day to day implementation, ministry of education for budgetary decisions). The program used a)pilots to adapt methodology to Ghana, b)larger scale RCT to test at scale and identify key constraints (tested four variations of remedial education delivery), c)explicit identification of implementation lessons through diverse steering committee that met periodically.	Yes [with caveat]. Constant exchanges between stakeholders occured to meet their respective needs: academics agreed to dramatic change in nature of intervention in midstream to adapt to operational realities of teachers. Ministry of education (and teachers) agreed to randomized nature of intervention. Timing decisions were made to adjust to political cycles. Project was officially led by the Ministry of Education for strategic and political reasons. But the political realities of electoral cycle were missed in scale up plan.	Mixed. The government's dominating question was the viability of the program at scale. To address this uncertainty, they needed to first determine whether approach could work within existing Ghanian constraints. Participatory mapping of learning pitches helped identify stakeholders' "currencies of exchange" and allowed for careful adaptation to local context. But ultimately did not identify political uncertainty for scale-up, adopting an appropriately hyper-local focus, yet sacrificing the broader, global view.
Graduating Ultra Poor (GUP) *Score: 4*	Yes [with caveat]. As part of international consortium, GUP was designed based on gaps of knowledge resulting from previous graduation interventions. Engaged local partners in definition of need to learn what worked to relieve poverty in Ghana. At the same time, did not engage local partners in their definition of success, which for all key local stakeholders included scale up.	Yes. Intervention was explicitly designed around existing knowledge and to systematically fill existing gaps. At the same time, while learning agenda was explicit at the international level, missed opportunity to include learning for local stakeholders on implementation, rollout, and scale up.	Mixed. IPA identified local expertise as critical and recruited a highly embedded NGO, which engaged local governments and community leaders to ensure adaptation to local needs. But stakeholders were subcontracted, not brought in as partners, missing opportunity for developing the capacity for scale up. In particular, did not grasp limits of RCT evidence vs. other types needed for buy-in (e.g. hand-off to local government for scale up).	Mixed. IPA was in charge of both implementation and evaluation. It incorporated voice of local stakeholders through NGO partner, changing the design of RCT in midstream. But while they identified key goals of local partners, they missed opportunities for broader impact and evidence integration, as focus was on appropriate execution of RCT, not on perspectives of local partners for scale-up and sustainability.	Mixed. GUP was rigorous about mapping out a measured sequence of experiments, however, failed to do so in an inclusive enough way— prioritizing the agenda of researchers and missing the priorities and expertise of local stakeholders. As a result, their learning pitches did not span the full breadth of uncertainty, nor did they integrate the necessary range of expertise in a distributed manner.

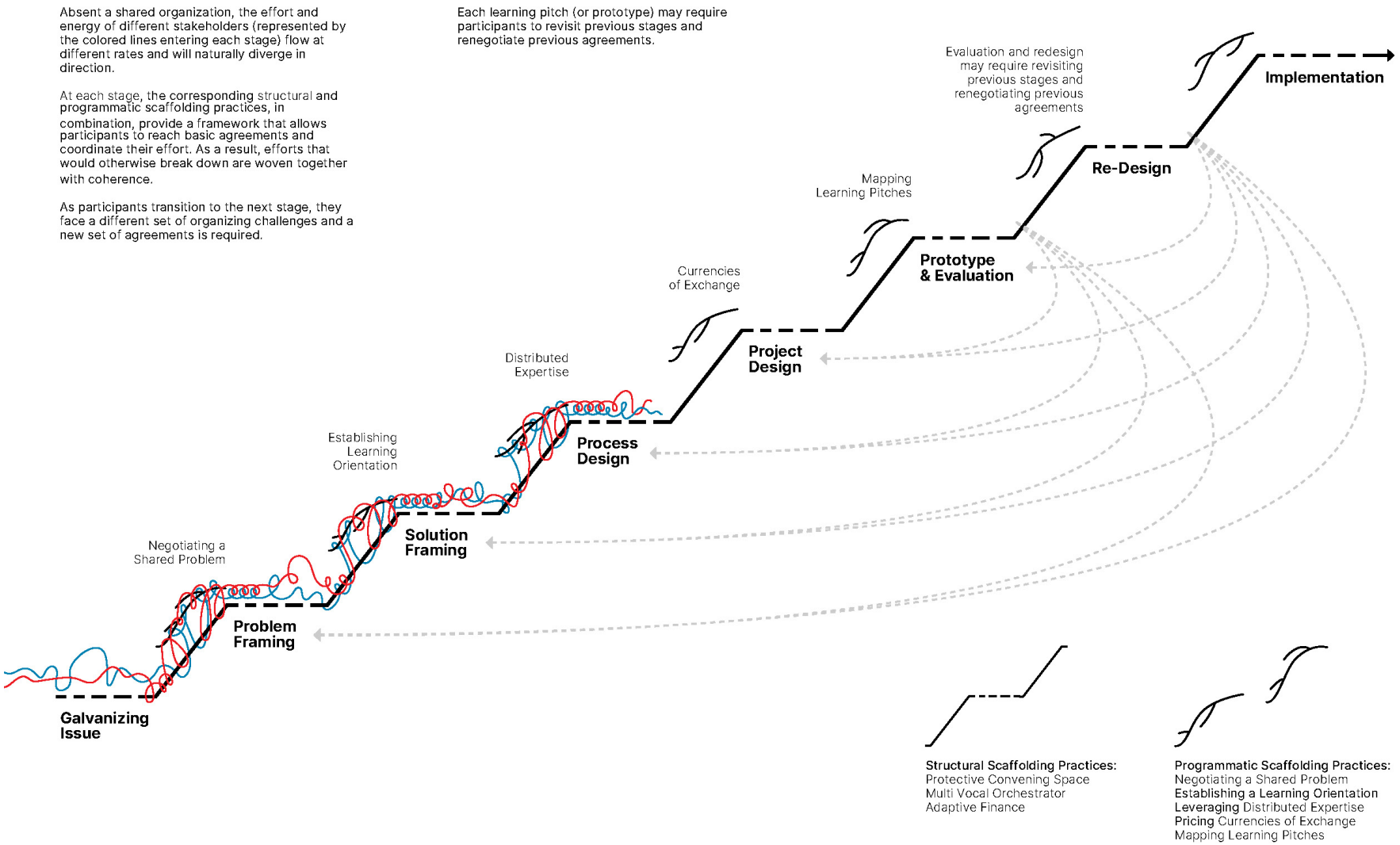
Table 6: (continued)

Name	*Negotiating a Shared Problem*	*Establishing a Learning-based Orientation*	*Leveraging Distributed Expertise*	*Pricing Currencies of Exchange*	*Mapping Learning Pitches*
FUEL *Score: 3*	No [but later yes]. Failed to identify the shared problem initially. Instead, went through several iterations to implement a top down vision of logistical improvement, without engaging downstream stakeholders. After several iterations (and three years), they gradually went down levels (from federal government to provincial government, then to local education officials, then to school principals), one iteration at a time, finally identifying a shared problem.	No [but later yes]. Focus was on implementation of vision of logistic optimization under assumption of government inefficiency. Founder and implementation team were extremely humble in face of failure, however, and adapted approach and framing dramatically to reframe objectives and take a radically different approach centered on adaptation to local school needs.	No [but later yes]. Focus was on using founder leverage with top government ministers to get "approval for implementation of vision" from the top. After failed attempts, the team recognized that it was not a logistics problem and that the most relevant expertise actually resided with local education officials and school principals.	No [but later yes]. Founding team initially thought they were the experts bringing their knowledge, assets, and money into the public sector to fix inefficiencies. Initially, local pushback was seen as evidence of lack of understanding and of evidence of bureaucratic red tape and corruption. Later, it was recognized that pushback was because problem had been misdiagnosed and focus shifted to currencies of exchange.	No. Struggled to appropriately frame the problem and prioritize key areas of uncertainty. Flexible funding allowed them to ultimately learn from failure, but under different circumstances their setbacks may have been fatal. E.g., they did not identify the program's dependency on the power structure of the education system and tested tools that missed basic field realities. This damaged morale and created important setbacks, which could have been easily avoided.
Aqua+ *Score: 2*	No. Assumed two key questions: 1)Effectiveness of the technology, tested in the lab. 2)Design of method to deliver it reliably, through product development. But distribution and take-up were not defined as part of the problem and were not incorporated. If individuals did not change their consumption habits, it must be because they did not understand the technology and product.	No. Focus was on efficient distribution and execution of business plan, as technology and product were assumed to be effective. Evidence from downstream retailers who had trouble stocking product and customers who were not changing consumption habits because product did not match their rhythms was systematically overlooked until very late in the process.	Mixed. Organization has culture of respect for basic science, product development, and design. But focus on technology and operations neglected downstream evidence, including on pricing and family rhythms, habits, and needs. E.g., product was piloted in slums close to headquarters, not where it was rolled out. Retailers were not engaged to understand their stocking needs.	No. Basic science and product development are separate units. Process structured as assembly line from basic science to product design to business plan execution. Downstream retailers and customers not identified as stakeholders with needs, assets, and constraints to factor into design. Funding by Antenna was potentially *too* flexible and not tied to evidence of local learning (goal was to learn across countries).	No. Aqua+ did not prioritize and then sequence learning. Focused on efficacy of technology without rigorously understanding user needs and motivations (i.e. assuming health was enough as motivation, neglecting convenience and accessibilty). Did not identify multiple areas of uncertainty, so insights that arose were obscured by focus on efficient execution.
Programa Primer Empleo (PPE) *Score: 1*	No. Focus was on generation of new jobs (results) without identifying underlying problems as defined and experienced by key stakeholders like employers or local governments.	No. Assumed that market rigidities from social security fees were a key constraint in hiring. Also assumed that businesses had little insight and would capture the program, that young workers were interested in the types of jobs created by the intervention, and that the subsidy would be enough for companies, etc. No assumptions were tested.	No. External stakeholders distrusted. Business associations sought to provide input but found no mechanism to do so. Evaluators' advice ignored because of political implications and budgetary restrictions. Focused on budget rules and proper execution of program, but left no room for integration of learning.	No. Subsidies were assumed to be attractive and sufficient. Program was designed to need minimal administration, with no mechanism to incorporate constraints, needs, or preferences of stakeholders or localities. Strict program rules constrained officials from adopting insights from internal evaluations.	No. Eligibility for the subsidy required companies to grow in employees. By not identifying the underlying macroeconomic uncertainity and the dependency of further design elements upon that condition, the program was derailed by economic downturn. Including employers in design process would have avoided this.

Table 7: Currencies of Exchange

Stakeholder	Time Cycles	Key Audiences	Typical Incentives	Resources	Non-negotiable needs	Evidence Priorities	Currency of Exchange
Funders	Typically annual funding cycles.	Board of directors, government regulators, the public (e.g. taxpayers, media, implementing partners, beneficiaries).	Ability to deploy money effectively. Alignment with organization's long-term strategy. Their own issue area exper-tise/credibility.	Access to funds. Access to diverse sources of situated knowledge (through grantees)	Compelling justification for use of funds. Avoidance of fraud or misuse of funds.	1.Their own expe-rience 2.Other fun-ders 3.Evidence gen-erated by their pro-grams. 4.Popular discussion (e.g. me-dia) 5.Academic re-search papers.	Willing to pro-vide capital in exchange for re-sults/recognition. Convening power. Support of cross-organizational exchanges (e.g. conferences).
Researchers	Peer-reviewed article publication timelines. Tenure timelines. Grant timelines.	Senior researchers making decisions on journal publications and academic tenure.	Publishing in peer-reviewed journals. Tenure. Awards. Fundraising.	Recognized fluency/mastery in separating complex problems into their constituent questions, framing hypotheses, identifying causality, and assessing other kinds of evidence.	Need to be able to publish and otherwise advance their careers in academia.	1. Their own evi-dence or other aca-demic evidence in their field published in peer-reviewed publications. 2. Popular research in their field (e.g. in media or elsewhere).	Ability to confer legitimacy through affirmation of being "evidence-based" in exchange expanded opportunities for ongoing access to data/information that furthers their research.
Intermediaries	Windows of opportunity for evidence translation, in which the timelines of other stakeholders align.	Policy makers. Implementers. Researchers.	Recognition and funding for being evidence-based.	Speaking multiple "languages," ability to translate ideas across stakeholder categories, convening power, serving as bridges across timelines.	Willingness of involved parties to engage in some compromise and make time for negotiation.	1. Evidence and themes popular among policymak-ers and funders. 2. Evidence with widespread im-plementation and replication poten-tial. 3. Published academic research	Willing to tap into networks, make markets for ideas, and translate between differ-ent stakeholder languages in ex-change for access to other stakeholders and influence in decision-making.
Policymakers	Typically annual government budget cycles. Electoral cycles.	Elected officials (for bureaucrats). Political party, media, public opinion (for elected officials).	Promotion and avoiding mistakes (for bureaucrats). Positive press and public opinion (for elected officials).	Ability to scale programs. Access to data. Convening power.	No negative optics/scandal. Abiding by legal mandate of institutions.	1. Their own experi-ence. 2. What other policy makers tell them. 3. Evidence from own country. 4. Evidence from other countries. 5. Peer-reviewed jour-nal articles.	Wiling to sup-port a program in exchange for affirmation of their policies. Conven-ing and mediation power/ability.
Implementers	Grant timelines. Budget cycles. Needs of beneficia-ries/clients.	Funders. Benefi-ciaries/clients.	On time and on-budget delivery of agreed upon projects. Low overhead costs.	Access to beneficia-ries/clients. Contextual sensitivity and intuition. Resourcefulness, resilience, agile problem-solving. Lessons and data from programs.	Enough resources to implement effectively.	1. Their own experience. 2. Pragmatic evidence (e.g., best practices on how to inte-grate/implement an idea). 3. Useful for fundraising (e.g., how can we design a project that is likely to be funded?).	Willing to provide access to benefi-ciaries/clients and related data in exchange for fund-ing and autonomy. Situated under-standing of how to apply knowledge. Agile adaptation to emerging challenges of implementation.

Figure 1: Process Model of Evidence Integration



References

Baldwin, Carliss Young, Kim B Clark, Kim B Clark, et al.. 2000. *Design rules: The power of modularity*, Volume 1. MIT press.

Banerjee, Abhijit V, Abhijit Banerjee, and Esther Duflo. 2011. *Poor economics: A radical rethinking of the way to fight global poverty*. Public Affairs.

Bechky, Beth A. 2003a. Object lessons: Workplace artifacts as representations of occupational jurisdiction. *American Journal of Sociology 109*(3), 720–752.

Bechky, Beth A. 2003b. Sharing meaning across occupational communities: The transformation of understanding on a production floor. *Organization science 14*(3), 312–330.

Bechky, Beth A. and Daisy E. Chung. 2018, Aug. Latitude or latent control? how occupational embeddedness and control shape emergent coordination. *Administrative Science Quarterly 63*(3), 607–636.

Benner, Mary J. and Michael L. Tushman. 2015. Reflections on the 2013 decade award –"exploitation, exploration, and process management: The productivity dilemma revisited" ten years later. *Academy of Management Review 40*(4), 497–514.

Brown, John Seely and Paul Duguid. 1991. Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organization science 2*(1), 40–57.

Brown, John Seely and Paul Duguid. 1998. Organizing knowledge. *California management review 40*(3), 90–111.

Brown, John Seely and Paul Duguid. 2001. Knowledge and organization: A social-practice perspective. *Organization science 12*(2), 198–213.

Brown, Tim. 2019. *Change by design, revised and updated: how design thinking transforms organizations and inspires innovation*. HarperCollins.

Brynjolfsson, Erik and Andrew McAfee. 2014. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Canales, Rodrigo. 2016. From ideals to institutions: Institutional entrepreneurship and the growth of mexican small business finance. *Organization Science 27*(6), 1548–1573.

Canales, Rodrigo and Jason Greenberg. 2015. A matter of (relational) style: Loan officer consistency and exchange continuity in microfinance. *Management Science 62*(4), 1202–1224.

Carlile, Paul R. 2002. A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization science 13*(4), 442–455.

Carlile, Paul R.. 2004, Oct. Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. *Organization Science 15*(5), 555–568.

Davis, Jason P. 2016. The group dynamics of interorganizational relationships: Collaborating with multiple partners in innovation ecosystems. *Administrative Science Quarterly 61*(4), 621–661.

Davis, Jason P and Kathleen M Eisenhardt. 2011. Rotating leadership and collaborative innovation: Recombination processes in symbiotic relationships. *Administrative Science Quarterly 56*(2), 159–201.

DiBenigno, Julia. 2018. Anchored personalization in managing goal conflict between professional groups: The case of us army mental health care. *Administrative Science Quarterly 63*(3), 526–569.

Dougherty, Deborah. 1992. Interpretive barriers to successful product innovation in large firms. *Organization science 3*(2), 179–202.

Evans, Joelle, Ruthanne Huising, and Susan S Silbey. 2016. Accounting for accounts: Crafting ethnographic validity through team ethnography. *Handbook of Qualitative Organizational Research: Innovative Pathways and Methods*, 143–155.

Faraj, Samer and Yan Xiao. 2006, Aug. Coordination in fast-response organizations. *Management Science 52*(8), 1155–1169.

Fayard, Anne-Laure, Ileana Stigliani, and Beth A Bechky. 2017. How nascent occupations construct a mandate: The case of service designers' ethos. *Administrative Science Quarterly 62*(2), 270–303.

George, Alexander L, Andrew Bennett, Sean M Lynn-Jones, and Steven E Miller. 2005. *Case studies and theory development in the social sciences*. MIT Press.

Gertler, Paul J, Sebastian Martinez, Patrick Premand, Laura B Rawlings, and Christel MJ Vermeersch. 2016. *Impact evaluation in practice* (Second Edition ed.). Washington, DC: The World Bank.

Gieryn, Thomas F.. 1983. Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review 48*(6), 781–795.

Glaser, Barney and Anselm Strauss. 1999. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Transaction.

Hausmann, Ricardo. 2016. The problem with evidence-based policies. *Project Syndicate 25*.

Howard-Grenville, Jennifer, Karen Golden-Biddle, Jennifer Irwin, and Jina Mao. 2011. Liminality as cultural process for cultural change. *Organization Science 22*(2), 522–539.

Kellogg, Katherine C.. 2009, 2019/10/03. Operating room: Relational spaces and microinstitutional change in surgery. *American Journal of Sociology 115*(3), 657–711.

Kellogg, Katherine C. 2011. *Challenging operations: Medical reform and resistance in surgery*. University of Chicago Press.

Kellogg, Katherine C., Wanda J. Orlikowski, and JoAnne Yates. 2006, Feb. Life in the trading zone: Structuring coordination across boundaries in postbureaucratic organizations. *Organization Science 17*(1), 22–44.

Leonardi, Paul M and Diane E Bailey. 2008. Transformational technologies and the creation of new work practices: Making implicit knowledge explicit in task-based offshoring. *MIS quarterly*, 411–436.

Lifshitz-Assaf, Hila. 2018. Dismantling knowledge boundaries at nasa: The critical role of professional identity in open innovation. *Administrative science quarterly 63*(4), 746–782.

Lingo, Elizabeth Long and Siobhán O'Mahony. 2010. Nexus work: Brokerage on creative projects. *Administrative Science Quarterly 55*(1), 47–81.

Mills, C. Wright. 1959, 09. The sociological imagination. *Journal of American History 46*(2), 355–356.

Obstfeld, David. 2005. Social networks, the tertius iungens orientation, and involvement in innovation. *Administrative science quarterly 50*(1), 100–130.

Obstfeld, David. 2012. Creative projects: A less routine approach toward getting new things done. *Organization Science 23*(6), 1571–1592.

O'Mahony, Siobhán and Beth A Bechky. 2008. Boundary organizations: Enabling collaboration among unexpected allies. *Administrative science quarterly 53*(3), 422–459.

Orlikowski, Wanda J. 2002. Knowing in practice: Enacting a collective capability in distributed organizing. *Organization science 13*(3), 249–273.

Pratt, Michael G., Douglas A. Lepisto, and Erik Dane. 2019. The hidden side of trust: Supporting and sustaining leaps of faith among firefighters. *Administrative Science Quarterly 64*(2), 398–434.

Puranam, P.. 2018. *The Microstructure of Organizations*. OUP Oxford.

Quintana, Chris, Brian J. Reiser, Elizabeth A. Davis, Joseph Krajcik, Eric Fretz, Ravit Golan Duncan, Eleni Kyza, Daniel Edelson, and Elliot Soloway. 2004, Jul. A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences 13*(3), 337–386.

Silbey, Susan S.. 2011. The sociological citizen: Pragmatic and relational regulation in law and organizations. *Regulation Governance 5*(1), 1–13.

Sorenson, Olav and Michelle Rogan. 2014. (when) do organizations have social capital? *Annual Review of Sociology 40*, 261–280.

Turco, Catherine J. 2016. *The conversational firm: Rethinking bureaucracy in the age of social media*. Columbia University Press.

Turner, Victor. 1969. Liminality and communitas. *The ritual process: Structure and anti-structure 94*(113), 125–30.

Valentine, Melissa A. 2018. Renegotiating spheres of obligation: The role of hierarchy in organizational learning. *Administrative Science Quarterly 63*(3), 570–606.

Valentine, Melissa A. and Amy C. Edmondson. 2015. Team scaffolds: How mesolevel structures enable role-based coordination in temporary groups. *Organization Science 26*(2), 405–422.

Weick, Karl E and Karlene H Roberts. 1993. Collective mind in organizations: Heedful interrelating on flight decks. *Administrative science quarterly*, 357–381.

Wood, David, Jerome S. Bruner, and Gail Ross. 1976. The role of tutoring in problem solving*. *Journal of Child Psychology and Psychiatry 17*(2), 89–100.

Zietsma, Charlene and Thomas B. Lawrence. 2010. Institutional work in the transformation of an organizational field: The interplay of boundary work and practice work. *Administrative Science Quarterly 55*(2), 189–221.

Zuckerman, Ezra W. 2010. Chapter 16 speaking with one voice: a "stanford school" approach to organizational hierarchy. In *Stanford's Organization Theory Renaissance, 1970–2000*, pp. 289–307. Emerald Group Publishing Limited.

Appendix. For Online Publication

Interview Protocol - Experts

NOTE: THE PROTOCOL SHIFTED SLIGHTLY FOR WITHIN-CASE INTERVIEWS, AS IT FOCUSED MORE ON THE SPECIFIC TRAJECTORY OF THE INTERVENTION, KEY DECISIONS TAKEN, ETC.

Consent

Thank you for agreeing to talk with me today. As a brief introduction, we are conducting a research project in partnership with the Hewlett Foundation to develop a better understanding of the process through which evidence is effectively (or ineffectively) incorporated into policy design, policy implementation, and field-level practices. Given your role and trajectory, we are interested in your personal perspective on how you typically make decisions, what types of evidence you have (or have not) incorporated into your decisions and why, and the constraints that you face in trying to inform your decisions with evidence.
Before we begin the interview, please know that

1. All information discussed is strictly confidential.
2. Your identity will never be revealed. If there are statements discussed through this interview that could identify you directly, they will not be used without your prior permission.
3. If, for specific research purposes, we would like to use a statement that could easily be traced back to you, we will contact you directly before using it to obtain your consent.
4. We would like to record this conversation, if it is all right with you. Our only intention is to retain our full attention on the conversation and not miss any nuances of your comments. If you are not comfortable with the audio recording, we can pause it at any point.
5. You should feel free not to answer any question. None of my questions should make you uncomfortable or should make you feel pressured to answer.
6. You may choose to stop the interview at any point in the process.
7. You may ask any questions or clarifications either about the process, the research, or the questions themselves at any point.
8. Please verbally acknowledge that you understand all of these points and that you agree with them.
9. Finally, is there a hard stop time that I need to keep in mind?

Interviewee History

While we have access to your CV and we know your professional trajectory, we always find it useful when interviewees describe their background, as it tells us a lot about the lens through which they view the world. In this context, we would like to start by asking about your current work

- How would you describe your current role?
- How would you describe the path that brought you here?

Organization

- What are the primary goals and mission of your organization?
- What kinds of relationships does your organization have with other organizations?
- How are they helpful, if at all?
- What are the biggest challenges you are facing as an organization today? What's the role of evidence in these?
- What are the biggest achievements you have attained as an organization? What's the role of evidence in these?

Evidence to Practice

Let's talk a little bit about this complex issue of incorporating evidence into practice. Based on your experience,

- What does it mean to you when I say "incorporate evidence into practice"? What specifically does that look like to you? (probe throughout the answer for each part of the statement... "what does evidence mean to you?" "what is incorporating?" "what is practice?") Can you give me an example?
- Has this definition changed for you during your experience? If so, how?
- Who are the other actors that you think would need to be involved, directly or indirectly, for this incorporation to happen? To what extent have they been involved in your processes and experience?
- What do you see as the (current and ideal) roles for each of those actors? How about the relation-

ships between them? (cooperation, animosity, competition, neglect)
- How similar do you think their definitions are of "incorporating, evidence, and practice"?
- Can you think of an example where you have seen this happen effectively?
- What do you think allowed for this to succeed here? (probe for role of each of the actors)
- Can you think of an example where it did not happen as effectively as you think it should have?
- What do you think were the main problems? And what were the sources of those problems? (probe for role of each of the actors)
- Are there factors that the examples (you just mentioned) have in common that you think contribute to successfully incorporating evidence into practice?
- When you think about the actors that need to be involved, what are factors that you think help them work towards the incorporation of evidence into practice? (if possible, change these words to better reflect the interviewee's definition of turning evidence into practice, probe for examples)
- What are factors that you think make it difficult for them to work towards that goal? (if possible, change these words to better reflect the interviewee's definition of turning evidence into practice, probe for examples)
- Thinking about your organization and your personal role, in what ways do you think that you are committed to the incorporation of evidence into practice? What allows you to do that effectively? What prevents you from doing it better?
- In what ways do you feel like you could be more committed? What would that entail?
- When you think about this issue in your field over the next ten years, what do you wish for?
- How is that different from today?
- What do you think are the main hurdles standing in the way?

Make sure to ask before time runs out:
- What else do you think I should have asked you? What questions would you ask if you were me?
- Is there anything that you might not have thought about before but that you realized during this interview?
- Is there anyone whom you think we must talk to?
- Is there anything you would like to ask me?