



Scaling Up Collaborative Social Accountability in Complex Governance Systems:

A Relational Approach for Evidencing Sustainability

Florencia Guerzovich and Alix Wadeson
JANUARY 2024



Abstract

When social accountability interventions scale up and their sustainability depends on the interactions of many agents and system components, related results are rarely observable at the end of an intervention. The 2019 OECD Development Assistance Committee's (OECD DAC) revamped evaluations criteria for assessing sustainability acknowledges that such results are often emergent, and should be monitored and evaluated with this in mind. It therefore emphasizes a turn towards assessing complex processes prospectively. It also asks evaluations to consider how likely it is that these results are evident at the time they are monitored or evaluated. However, the social accountability field continues to have gaps regarding doing this effectively in practice.

This paper presents and provides evidence from testing an innovative operational approach that has promising potential to support this aim - a sequential, relational rubric. This approach can support practitioners to monitor, evaluate and learn about the causal processes of scale up of social accountability interventions with an eye towards sustainability i.e., considering prospective sustainability. It is grounded in systems thinking, co-production and social learning theory, as well as links with collective governance and social contract theory for development.

Evidence yielded from the authors' testing of this approach on a sample of diverse projects from the Global Partnership for Social Accountability (GPSA) program revealed that the alleged 'absence of evidence' dilemma of social accountability scale up is due to ill-fitting concepts and methods for assessment. It challenges existing assumptions and findings that claim that social accountability processes do not scale and are unsustainable. The authors propose that by using fit-for-purpose concepts and methods with a focus on social learning and compromise - also called a 'resonance pathway to scale' which this paper discusses in detail - it is possible to observe loosely coordinated scale up processes at work in many (but not all) social accountability interventions and identify tangible evidence of prospective sustainability. An important caveat is that these processes, the outcomes they generate, and the corresponding evidence often look qualitatively different than the original intervention design and predictions for scale-up at that point in time. This is because the process of deliberation and compromise inherent to social accountability work in dynamic local systems introduces changes and new conditions for uptake by diverse actors in the public sector, civil society, and donor institutions.

The paper concludes that even relatively small-scale localized projects of three to five years with budgets of less than one million USD, across different contexts and sectors can produce processes and outcomes which contribute to many forms of sustainability, including via scale-up. Furthermore, the cross-fertilization of learning and aggregation of results for scale-up across projects within and beyond the GPSA (and other programs) can help monitoring evaluation and learning (MEL) and social accountability practitioners alike to deliver on a program's mandate. Doing so can also create new knowledge for the wider social accountability field that siloed interventions, lacking suitable concepts and methods for assessing scale-up and prospective sustainability, often fail to produce. The paper ends with recommendations for taking forward this approach and the associated benefits, implications and required investments.

Acknowledgments

This paper represents a culmination of evolving learning, evidence, research and practical experience from the authors and the wider field of social accountability. A key source of findings and evidence base for the paper is from our respective work with the World Bank's GPSA programming over the past decade, in partnership with civil society, the public sector, communities and citizens around the world. This paper reflects the contributions of numerous stakeholders engaged with the GPSA, especially Jeff Thindwa, Ann-Sofie Jespersen, and Aly Zulficar Rahim. We greatly appreciate the valuable feedback and time provided by peer reviewers Mathieu Cloutier and Tom Aston. We also acknowledge the work of other social accountability practitioners, researchers and evaluators for their contributions to social accountability work and evidence building for the wider field. This paper builds upon the existing evidence base to support improvements and offers insights about how social accountability programming can be strengthened, sustained and scaled.

Copy Editor: Amber Meikle

Designer: Mohamed Elmahdy

Cover Image: © Curt Carnemark / World Bank. Further permission required for reuse

Suggested Citation

Guergovich, Florencia and Wadeson, Alix. 2024. Scaling up Social Accountability in Complex Governance Systems: A Relational Approach for Evidencing Sustainability. World Bank, Washington, DC.

Contact Information

For questions and other information about this paper and its findings, please contact the authors: alixwadesonconsulting@gmail.com and florcig@gmail.com

© 2024 International Bank for Reconstruction and Development / The World Bank 1818 H Street NW | Washington DC 20433

Telephone: 202-473-1000

www.worldbank.org

Glossary of Key Concepts

Accountability: A social relationship between a power-holder, an actor that performs a task such as a government official, and an account-holder, those for whom the task is performed or who are affected by it. In this social relationship the power-holder is, in practice, obliged to be more transparent and explain and justify their decisions, behaviors, trajectories, and results (answerability), with information and transparency. There is the possibility of dialogue among the parties and the account-holder can pose questions and ask for rectifications, remedies, corrective action or problem solving (accountability processes). As a result, the account-holder can pass judgment and the power-holder can face some form of consequences. These consequences can be hard or soft, formal, and explicit or informal and implicit, sanctions in the case of malperformance. In other words, accountability entails a proactive set of processes and practices where the how – the quality of the social relationship between a power holder and an account holder – is the what (Guerzovich, 2022, drawing on Bovens et al, 2014).

Social Accountability: Processes that seek to make communities leading agents in their localized development story by: (1) improving the quality of goods and services, (2) primarily through monitoring and oversight of those goods and services, (3) citizens' collective, rather than individual, efforts to hold power-holders (primarily service providers and bureaucrats, secondarily politicians) to account, (4) providing a concrete mechanism to rework the social contract and strengthen local systems, in the sense of transforming state-society relationships and the norms and power dynamics associated with them (Guerzovich and Aston, 2023).

Collaborative Social Accountability: Processes whereby civil society organizations and public sector institutions with decision-making power and public management authority at different levels across the institutional and service delivery chain convene to analyze a problem, identify citizen participation mechanisms to help solve it, and agree on joint actions to co-produce solutions and appropriate responses (Poli and Guerzovich, 2020). This is a term coined by and applied in all of the GPSA's programming.

Learning: The process of creating new knowledge, insights, or understanding – usually about what works, what might work, or

what doesn't work in advancing a given goal. This paper, and the resonance pathway to scale, is most interested in shared learning – learning that happens with others, also known as joint learning or social learning (Guerzovich et al, 2022 drawing on Wenger-Trayner and Wenger-Trayner, 2021).

Resonance Pathway to Scale: This expects social accountability to scale up based on deliberation, compromise, and coordinated collective action among diverse actors (Guerzovich et al 2022). The logic is that social accountability processes contribute to overcoming the challenges of collective action in a game theoretical sense (Ostrom, 1990; also see World Bank, 2017). Its main thrust is social learning. That means enabling a group of individuals to organize and work out how to make the most of a situation (e.g., insights learned by implementing social accountability in select locations) to create shared gains (e.g., using those insights to inform decisions in other locations) through loose coordination and collaboration.

Scale/scale-up: The ability of a project or program to grow its effects beyond its sectoral and geographic boundaries, to reach more people (Guerzovich and Poli, 2014).

Sustainability: When and how a project's net benefits continue or be likely to continue after the end of the project (OECD DAC, 2019).

System: The interconnected set of factors (policies, practices, resource flows, relationships and connections, power dynamics and mental models) that jointly produce a development outcome – the whole is greater than the sum of its parts (Kania et al, 2018).

Uptake: In the context of the GPSA and this paper, these are actions taken by public service sector actors, policymakers, practitioners, and other development actors that facilitate and contribute to the adaptation, application, and/or sustainability of elements of collaborative social accountability processes (e.g., approaches, strategies, tools, mechanisms) and/or the application of lessons and insights from collaborative social accountability programming and evidence. This definition of uptake includes many types of sustainability – actual, prospective and scale-up.

Executive Summary

The Paper at a Glance

In 2019, the Organization for Economic Co-operation and Development -Development Assistance Committee (OECD DAC) revamped its criteria to assess interventions. This new criteria puts front and center the methodological challenges associated with evaluating sustainability. Sustainability is defined as *“when and how a project’s net benefits continue or are likely to continue after the end of a project”* (OECD DAC, 2019). When sustainability results are dependent on the interactions of multiple actors and elements in a complex local system, they are often emergent over time, prospective in nature, and uncertain at the point of a project’s final evaluation.

This paper presents an innovative operational approach – a sequential relational rubric – to monitor, evaluate and learn about the causal processes of scale-up, with an eye towards sustainability (i.e. considering prospective sustainability). Scale up – *“the ability of a project or program to grow its effects beyond its sectoral and geographical boundaries, to reach more people”* (Guerzovich and Poli, 2014) – and sustainability are not synonymous. However, in the social accountability field, many assessments consider scale-up as an essential pathway and key indicator of project sustainability. Accordingly, much of the relevant reviews and literature in the field have found that while many projects achieve some form of positive results, there is limited tangible evidence to demonstrate they have successfully scaled-up to reach more locations and people (see for example E-Pact Consortium, 2016). This commonly held conclusion infers that such projects are not sustainable, contributing to a pessimistic narrative about the potential and long-term impact of social accountability programming overall (Aston, 2021).

Written for fellow monitoring evaluation and learning (MEL) practitioners working in the social accountability space, this paper argues that absence of evidence of sustainability to date is not evidence of absence in practice, nor does it equate to social accountability projects being unsustainable. Rather there is an evidence challenge, which lies in the ill-fitting concepts and methods that are often used to monitor and evaluate scale-up of social accountability projects. The authors propose that by applying fit-for-purpose concepts and methods that focus on social learning and compromise, it is possible to observe processes of movement towards scale-up. These different forms of prospective sustainability can be evidenced in a significant proportion of social accountability projects. The relational rubric presented and discussed herein builds upon and strengthens the recent theoretical proposition of the resonance pathway to scale (Guerzovich et al, 2022). This pathway asserts that the scale-up of many social accountability processes involves social learning at its core; and that such processes may occur gradually based on deliberation, compromise, and coordinated collective action among diverse actors. An important caveat to note is that these processes, the outcomes they generate, and the corresponding evidence often look qualitatively different from the design and predictions for scale-up of the original intervention. This is due to changes and conditions for uptake that emerge, both throughout a social accountability project and beyond its implementation.

The approach presented and evidenced in this paper can support practitioners to monitor, evaluate and learn about the causal processes of scale-up of social accountability interventions with an eye towards sustainability. It is grounded in systems thinking, co-production and social learning theory, and links with collective governance and social contract theory for development. All these models also underscore the uncertainty and emergent nature of complex and relational processes, validating the need for conceptual and methodological approaches that sufficiency account for such dynamics.

Accordingly, the paper discusses new empirical findings and a wealth of examples yielded through the authors’ test of this sequential relational rubric across a sample of 15 completed projects directly supported by the World Bank’s [Global Partnership for Social Accountability \(GPSA\)](#). The GPSA’s approach to MEL at a portfolio (or program) level enabled the iterative development of this rubric method through quick feedback cycles of learning and adaptation. The paper concludes that even relatively small-scale localized projects of

three to five years with budgets of less than one million USD, working across different contexts and sectors can produce processes and outcomes which contribute to many forms of sustainability, including via scale-up. Furthermore, the cross-fertilization of learning and aggregation of results for scale-up across projects within and beyond the GPSA (and other programs) can help MEL and social accountability practitioners alike to deliver on a program's mandate. Doing so can also create new knowledge where *“the whole is greater than the sum of its parts”* (Guerzovich, 2021b). With further testing of the rubric approach and building the evidence base for the resonance pathway to scale, the paper proposes that a solution for the absence of evidence dilemma is possible. If relevant organizations and funders commit to and invest sufficiently in portfolio-level MEL that is grounded in fit-for-purpose concepts and methods for assessing scale-up, then the narrative can shift towards prospective sustainability in its many forms, recognizing the promise of long-term impact from social accountability programming.

More Meaningful Monitoring and Evaluation of Relational Social Accountability Processes within Dynamic Local Systems

The sequential relational rubric for assessing scale-up and other forms of sustainability, and the concepts embedded within it, are aligned with and help to operationalize the revised OECD DAC evaluation criteria for sustainability (Guerzovich, 2023 and Guerzovich, 2023a). Key features of the concepts and criteria applied in the authors' work, and discussed in the paper are:

- **A systemic lens that considers how interventions fit within local systems and effect different actors.** A systemic lens focuses on interactions of a wide range of actors in a system, rather than a narrow focus on the siloed actions of the project's direct civil society implementers. This is essential to capture scale-up processes because they often rely on downstream actions taken by others in the systems. These include public sector institutions, funders, and other development agencies that adopt, adapt and/or sustain elements of social accountability processes in different ways after a project ends.
- **An emphasis on prospective sustainability.** During a project's life and at its closure, it is not possible to have certainty about the future in a complex system (or any system). However, it is both possible and desirable to focus on the likelihood of sustainability, and scale-up as one form of it. Therefore, assessments should be based on signals for prospective sustainability and uptake, given uncertainty using both monitoring data from the whole project life, and triangulated evidence at the final evaluation stage. Emphasis on both forms of data is critical because ongoing attention to monitoring data helps project teams to identify, plan for and build opportunities that can support the continuation of positive effects of a project, from the point of its design, while also mitigating barriers and risks along the way.
- **A focus on function over form.** Social accountability processes and their scale-up will vary widely in their form i.e., the tools, strategies and mechanisms selected for implementing social accountability work. The variance between contexts considers different perspectives, relationships, and incentives of the key actors who can drive scale-up in the long-term. Therefore, the forms of social accountability processes need to be localized to different systems of implementation. At the same time, they often play a similar function: improving public service delivery in a collaborative manner that includes communities and citizens. Discrete elements (or components) of a social accountability process will often be adapted and applied in many forms, rather than the whole process replicated completely. When evaluations look only for complete replication of a process that is the same as its original design, it fails to capture other forms of scale. This is not only an unrealistic expectation for social accountability projects, but also discounts ways that different actors can engage in more meaningful and responsive social accountability relationships, while also contributing to sustainable outcomes and scale.

These features are often lacking in traditional monitoring and evaluation processes, which the paper discusses in greater technical detail with examples. However, applying them challenges the (erroneous) conclusion that if we cannot evidence or demonstrate the lasting change during or right at the end of a project, then movement towards scale-up or other forms of sustainability are not happening and will not continue. When uncertainty is rife and insufficient time has elapsed to observe sustainability at work, the authors propose that evaluators should apply concepts and methods that can assess the conditions required for actual sustainability and the likelihood of prospective sustainability (rather than the certainty), in a wide range of different forms.

Defining and Evidencing ‘Good Enough’ Results and ‘What Counts’

Most of the existing literature on social accountability sustainability and scale-up focuses on wholesale replication or complete institutionalization of social accountability processes, equating such outcomes with success. It thus fails to capture legitimate outcomes that include adaptation, incremental progress, fits and starts, and gray areas (for alternative approaches, see Integrity Action, 2020). Instead, the resonance pathway to scale and the relational rubric approach for assessment recognize that success depends heavily on the interactions between and actions of several actors within a given system and considers how these dynamics evolve over time to yield positive results for sustainability. But what does this look like in practice, and what counts? And how can we sufficiently evidence it with a MEL system and its data?

While acknowledging that there are no perfect definitions and that concepts change with learning and practice, this paper argues that striving for ‘good enough’ is reasonable, while also being careful about conceptual stretching -- defined as *“the distortion that occurs when a concept does not fit the new cases”* (Collier and Mahon, 1993). The approach presented is based on the assertion from the GPSA that: a result is demonstrated when lessons from or elements of collaborative social accountability inform decisions and actions taken by the public sector and other civil society and development actors beyond an individual project, including after the project has ended. Such results are often associated with the uptake of selected element(s) of a collaborative social accountability process, rather than wholesale scale-up of it.

The relational rubric method was developed and then tested through the assessment of the associated operational indicator in the GPSA Results Framework, applied across its portfolio of projects:

“The percentage of GPSA grants in which public sector institutions and other relevant actors (e.g., the World Bank, other donors, civil society organizations) seek to:

- i. use substantive lessons for improvements of targeted policies, processes, and mechanisms;*
- ii. apply or sustain elements of collaborative social accountability processes after life of the project;*
- iii. adapt insights from GPSA projects to scale them through programs or policies; or*
- iv. apply elements of collaborative social accountability processes in additional localities or sectors.”*

It is important to emphasize that the ‘seek to’ part of the indicator statement is critical because the uptake of GPSA projects is contingent, in that it can be introduced but not sustained by the project after its closure.

Therefore, the authors’ assessment of this GPSA result and indicator casts a wide net for ‘what counts’ for sustainability, including examples of scale-up. Relevant examples evidenced across the GPSA portfolio, and its respective projects include:

- Work by key stakeholders closely engaged directly in a project (i.e., public sector officials, GPSA/ World Bank project personnel, and representatives from civil society and community organizations and networks integrated into another public sector project or program.
- Public sector counterparts used lessons to inform public sector reforms and policies.
- Emulation by local public sector or service providers (e.g., education officials and schools) that observed, adopted, or adapted the collaborative social accountability process from a project.
- The World Bank or other funders used lessons and approaches to advise public sector or other development partners’ programs.
- The World Bank or other funders financed an adaptation of the project in the same or other sectors.
- Any observed or reported uptake, sustainability and/or scale-up led by other international non-government organizations (INGOs) or civil society organizations (CSOs).
- The project actions and trajectory demonstrated ongoing dialogue with key actors (relevant public sector officials and World Bank operations staff) to move the process for potential uptake of collaborative social accountability processes forward.

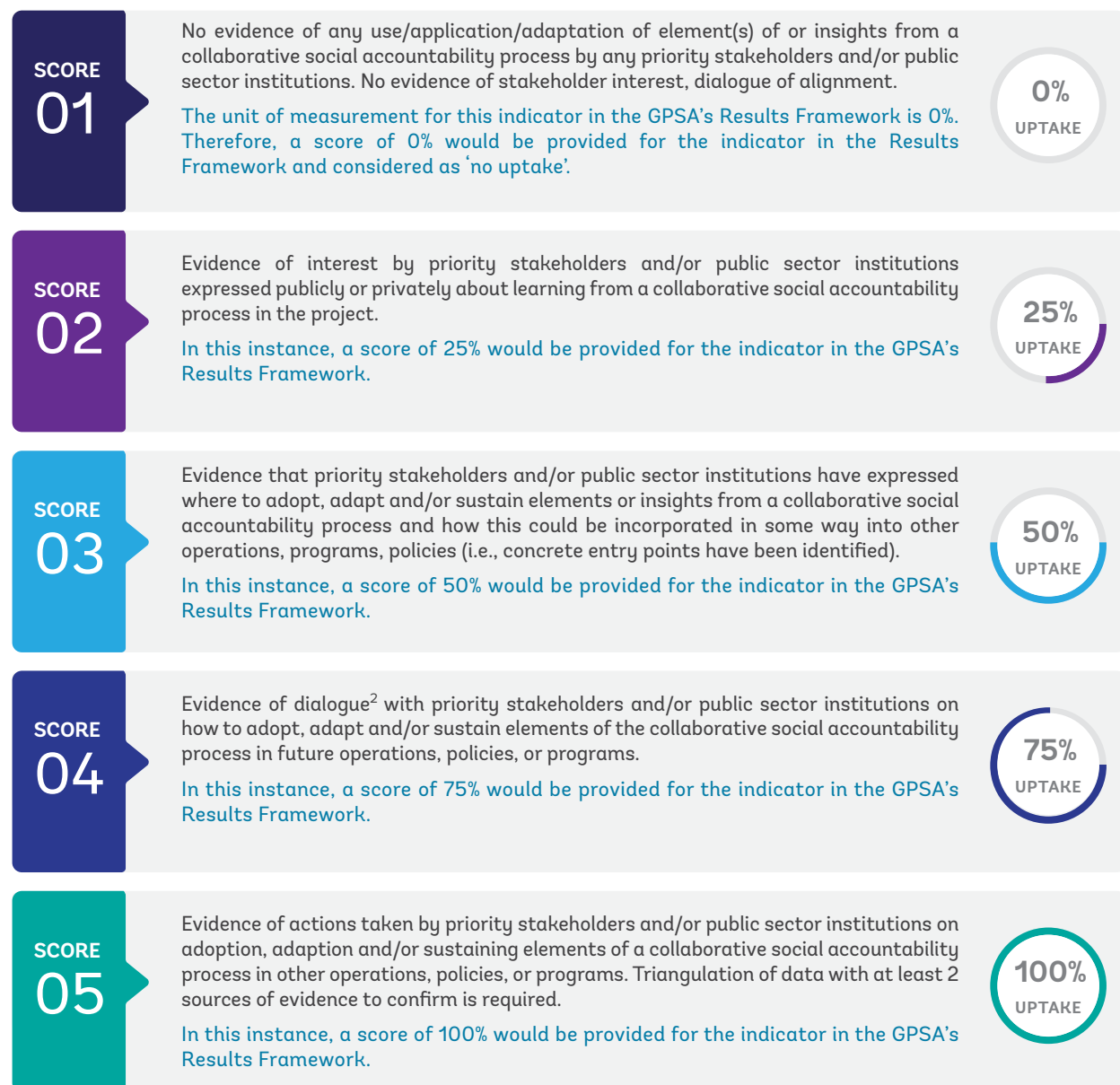
Understanding and Incorporating a Causal Sequence

A critical part of evidencing the likelihood for scale-up and prospective sustainability is to first understand and then investigate the concrete and sequential steps involved in these processes. This relational rubric also has an innovative sequential component. It organizes relevant actions and events in a temporal order to help identify if and how scale-up is on the right track or not, with an eye towards prospective sustainability. Such sequencing can provide significant leverage and support for project teams and evaluators to causally trace complex change processes and produce plausible explanations when concrete outcomes are still unknown. The rubric uses a five-point scale based on these sequential steps, with respective criteria for each level, moving from none to partial to full uptake of collaborative social accountability processes.

In recognizing the reasonable limits and appropriate expectations for sustainability and scale, a score of 5 or 100% does not equate to wholesale uptake or replication, in the context of this rubric, for the many reasons discussed above.¹ This interpretation may be different to the use of percentages by other MEL practitioners and assessment methods. However, translating each level (1-5) in the rubric with a percentage score provides a common reference point and metric that is easily comparable and transferable across different projects and

¹ As per the GPSA’s evolving Theory of Action and by design, GPSA projects do not intend or expect to achieve wholesale uptake of a collaborative social accountability process within a given sector and country of operations, given the limited budget and time-frame and their experimental nature. This aligns with the authors’ conceptions and evidence about what is realistic to expect for sustainability and scale-up of social accountability programming operating in complex governance systems of intersecting and continuously shifting political, economic and social dynamics of influence.

programs (for more detail on GPSA outcomes, indicators and application of the rubric, see the [MERL Guide for GPSA Grant Partners and Consultants](#)). Use of or adaptations of this rubric can eliminate percentages and adjust criteria for different levels in the scale as long as the core features of reasonable expectations ('good enough'), sequential causal steps, and transferable units of measurement are still applied.

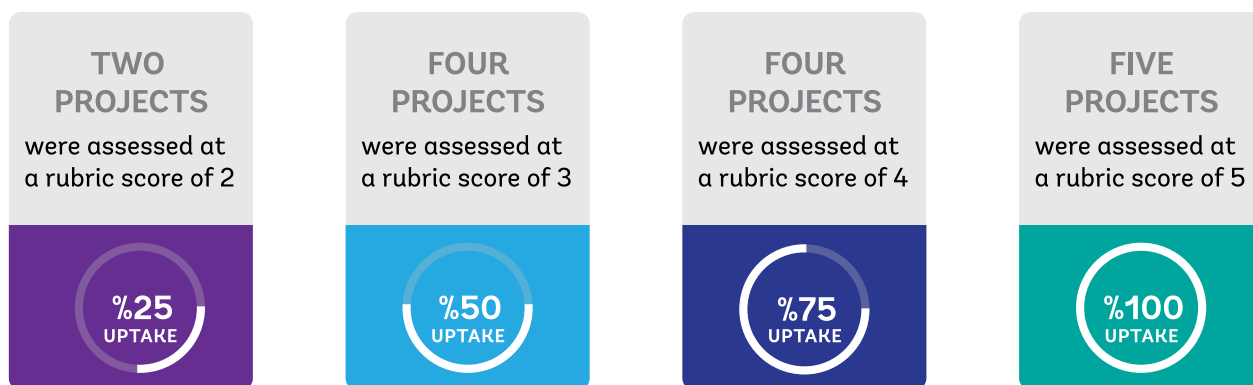


Source: Adapted from Wadeson and Guertzovich, 2023

² The dialogue with priority stakeholders is done by the project (e.g., the project team, grant partners).

Empirical Findings and Benefits to Date

After developing this relational sequential rubric, the authors tested it by systematically assessing an existing dataset (e.g., project reports, knowledge products, independent evaluations, official Implementation Completion Reports), from a sample of 15 closed GPSA projects, against the indicator presented above. This exercise identified numerous examples of scale-up, actual, and prospective sustainability in various forms from the source evidence. The breakdown of rubric scores and associated percentages from the assessed sample are:



Source: World Bank data

The findings validated the promise of the resonance pathway to scale as well as the feasibility and applicability of the relational rubric approach to evidencing it, despite the experimental nature of the relational rubric and the limitations of the exercise. The testing enabled adaptive learning and improvements were made to the rubric.

The findings challenge the 'absence of evidence' dilemma regarding the sustainability and scale of social accountability work. They demonstrate that it is plausible for relatively short projects of three to five years with budgets of less than one million USD to contribute to actions taken by the public sector and other priority stakeholders to adopt, adapt and/or sustain elements of a collaborative social accountability process in other operations, policies, or programs, moving along a resonance pathway to scale. However, in line with the key features and concepts embedded in this approach, the forms found across projects were diverse and did not look the same, but still had similar functions. This reflects and supports the central notion that context-specific processes of interactions, deliberation, social learning, compromises, and loosely coordinated collective action manifest in various ways (forms), yet they are still coherent with and strengthen local dynamic systems and processes for improved public service delivery and policy (functions).

Another benefit yielded through applying this relational rubric method is its potential to help move beyond a siloed understanding of projects within a program or portfolio; emergent findings can foster synergies and cross-learning between projects and aggregate results at a higher level. In this case, the rubric's application for each sample project and the aggregation of these results at the portfolio level, produced knowledge in a way that is not possible through evaluating projects individually without a transferable method. This added value contributes to the GPSA's corporate mandate and offers a practical means for other social accountability programs (or related fields) to do the same.

Recommendations

With more testing and iteration of the relational rubric, and intentional design and implementation of projects and MEL systems to apply it in real-time with primary data, the authors propose that the evidence base for the resonance pathway to scale will grow. This can in turn create new knowledge and influence a shift in discourse about the potential long-term impact of social accountability work, challenging the 'absence of evidence' dilemma in the field.

To meet these interlinked aims, the authors recommend that:

- » The GPSA and other funders and organizations working on social accountability make intentional and long-term investments in robust research and evaluation initiatives using the relational rubric to assess actual and prospective sustainability and scale-up, based on the key concepts, features and preliminary evidence presented in this paper.
- » Ensuring that funders, practitioners and evaluators hold realistic expectations for the success of collaborative social accountability processes and individual projects, recognizing the many legitimate forms of sustainability, the incremental steps involved, and long-term time-frames required for scale-up.
- » Targeted selection criteria and planned sufficient resourcing of external evaluators and internal MEL staff with the appropriate skillsets to monitor and evaluate programming in this way.
- » A supportive leadership environment and sufficient investment for systematic assessments of scale-up and sustainability at project and program levels, repeated consistently over time.

01

—

Introduction

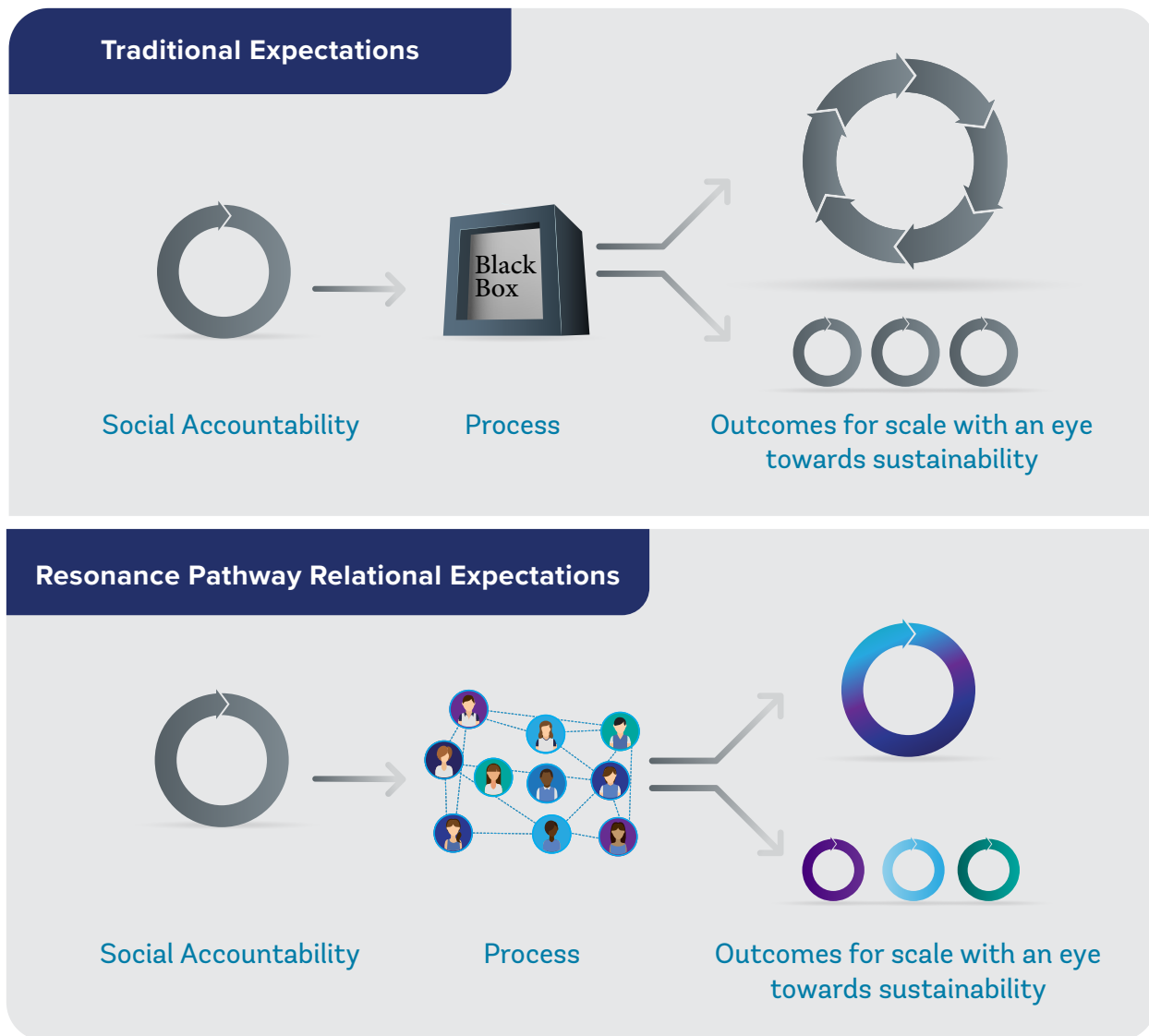
Scale (or scale-up) is *“the ability of a project or program to grow its effects beyond its sectoral and geographical boundaries, to reach more people”* (Guerzovich and Poli, 2014). It is often referred to as the Achilles’ heel of social accountability programming and various other participatory approaches to development. Relevant reviews and literature in the social accountability field find that while many interventions achieve some form of positive results, there is limited tangible evidence to demonstrate they have successfully scaled-up to reach more locations and people (E-Pact, 2016; Fox, 2014).

Sustainability in the simplest terms is defined as *“when and how a project’s net benefits continue or are likely to continue after the end of a project”* (OECD DAC, 2019). While a social accountability project can be sustainable by yielding net benefits that continue beyond the project life without scaling-up (Guerzovich and Poli, 2014), many social accountability assessments consider scale-up as a key pathway and indicator of project sustainability (see Guerzovich, 2022c).

While scale and sustainability are not synonymous terms, the prevailing assumption in the social accountability literature is that scale-up is required to evidence sustainability. The failure to scale-up by the end of an intervention is often assumed to suggest a lack of sustainability in the future, contributing to a pessimistic narrative about the potential and long-term impact of these interventions and broader social accountability work (Aston, 2021). These pessimistic and ill-fitting assessments of the potential of social accountability processes conceptualize and evaluate scale with an eye towards sustainability (i.e., considering prospective sustainability) in simplistic and unrealistic terms. The traditional evaluation approach found in the current evidence base often reflects scenarios in which social accountability processes will look qualitatively the same from their design to the end of an intervention, and the expected changes are expressed quantitatively.

The first diagram in Figure 1 illustrates this assumption as a ‘scale-up transmission belt’ by showing a gray ring that enters a black box and then produces bigger or more replicated gray rings. Yet, when the changes interventions seeks to make are complex and contingent on many others actors in a system who bring with them their own circumstances and agendas, what happens inside the black box is critical to informing the expectations for and assessment of results. Inside the black box, the causal path towards scale-up is rarely linear while the results to which they contribute are diverse, as illustrated in the second diagram in Figure 1: Comparing Expectations of Outcomes for Social Accountability Scale-up in Complex Governance Systems. To adequately understand and assess this phenomena requires a different approach to gauge whether projects are on the right track towards meeting their goals for complex change, and the potential for benefits to continue and evolve in the longer-term (Haldrup, 2020).

Figure 1: Comparing Expectations of Outcomes for Social Accountability Scale-up in Complex Governance Systems



Source: Own elaboration

This paper is aligned with and builds upon other salient factors emerging in both evaluation and social accountability fields: the Organization for Economic Co-operation and Development - Development Assistance Committee’s (OECD DAC) revamped evaluation criteria to assess sustainability of interventions (OECD DAC, 2019); the lively debate about how to connect (Patton, 2020) and apply (Kania et al, 2018) this within a systems thinking lens; and literature about co-production and social learning (Ostrom, 1990; Doin et al 2012; Wenger-Trayner and Wenger-Trayner, 2021). These all provide useful stepping stones for re-thinking, defining and developing fit-for-purpose concepts and methods for evaluating complex change processes both within and beyond the social accountability space. However, there is a dearth of operational and transferable approaches to put this cutting-edge thinking into practice.

To help address the challenges of monitoring and evaluating the scale-up and sustainability of complex change processes both conceptually and operationally, this paper presents an innovative method developed for this purpose - a relational rubric. Using the evidence from applying this method to a sample of social

accountability projects, the paper explores whether and how social accountability processes and the emerging insights support scale-up during and beyond a project life; how these results can be better evidenced through individual projects; and how they can be aggregated upwards to provide more comprehensive knowledge about the sustainability of social accountability programming.

The theoretical underpinnings of this relational rubric is the resonance pathway to scale, which has been ignored or missed by most research and evaluations in the social accountability field.³ As will be discussed further, the resonance pathway opens up the black box and identifies the many forms of results to which localized social accountability processes contribute. It also accounts for what can happen when stakeholders that were not involved in their design and/or implementation of such processes, encounter learning and insights that emerge from them. Accordingly, this paper conceptualizes scale-up with an eye towards sustainability as a gradual, sequential process of joint or social learning, deliberation, and compromise. All these interactions are central to support, trigger, and contribute to scale-up in ways that support adaptive uptake. However, these often look like loosely coordinated collective action rather than replication of an original model in which all forms of scale look qualitatively the same.

See Box 1 for two scenarios that illustrate these dynamics at play with examples of qualitatively diverse emergent outcomes that can result through a resonance pathway to scale, as also depicted in Figure 1.

Box 1: Scenarios Illustrating the Emergent Outcomes That Can Result through a Resonance Pathway to Scale

Scenario 1: Actors such as the mayor of a different village, the person responsible for an education region, a bureaucrat in a national ministry and a staffer in a different donor agency are not directly implementing the social accountability process of a given civil society-led intervention. They may have awareness or some engagement at different points. They will often look at the lessons emerging from the implementation of a social accountability process. They will also consider what value it can add to the work they are doing in different locations or sectors in the system. Due to their different roles and interests within the local system, they will bring in new perspectives and considerations. Therefore if and when they decide to scale up the social accountability process, their version of the process and its scale-up will often be an adaptation of the original design; this is an emergent outcome rather than one that could be accurately anticipated at design. The resulting process will take on new properties, with some components taken up, and others changed or dropped. This happens as part of the deliberation and compromises needed to enable the scale-up for sustainability of a given social accountability process.

Scenario 2: A civil society group implementing a social accountability process finds out that a ministry may incorporate parts of its process, but will only accept to focus on some aspects and integrate some new protocols. This would be a form of scale-up which could reach many locations because of the ministry's role within the system. An alternative would be to multiply facilitators that can effectively implement a social accountability process. From a handful of committed and experienced professionals hired by a civil society group to hundreds, hired by another public or civic organization to reach new locations. This would be a huge task, and without compromise on the process between the designers and implementers (the civil society group and professionals) and other actors who can facilitate scale-up, it is unlikely that the ministry would choose to scale-up the social accountability process.⁵

Source: Own elaboration

3 The resonance pathway to scale introduces two important theoretical innovations. First, a pathway to scale that reflects the lived experience of many practitioners but has been overlooked by traditional schools of thought in the field, which prioritized best practices and resistance as the driving forces for scale up and sustainability. Second, traditional schools of thought have often presented their preferred pathway to change as universally applicable, despite mixed results. The resonance pathway to scale focuses on the conditions under which it and other models may be better bets and their limits (Guerzovich et al, 2022). Also see Aston, 2022.

5 Thank you to Thomas Aston for sharing this example from a GPSA project.

Moving beyond conceptions to practice, the authors systematically applied the relational rubric method on a sample of 15 closed projects directly supported by the World Bank's [Global Partnership for Social Accountability](#) (GPSA). They used existing project documentation (e.g., project reports, knowledge products, independent evaluations, Implementation Completion Reports), as well as GPSA staff's tacit knowledge and accounts to triangulate the findings. A rubric assessment was determined for each project and then aggregated across the sample. This new dataset identified numerous examples of scale-up, actual, and prospective sustainability in various forms. The resulting evidence helped to validate the resonance pathway. Learning from the process supported testing and improvement of the rubric.

The exercise revealed that absence of evidence about social accountability scale-up to date does not equate to evidence of absence in practice, nor does it equate to social accountability processes being unsustainable. Rather, the core problem is one of ill-fitting concepts and methods for assessment of social accountability processes and outcomes. Such processes and outcomes may be familiar to practitioners but are only observable and evidenced when monitoring and evaluation of interventions captures systemic processes at work between different actors, as the relational rubric does. This relational rubric also has an innovative sequential component. It organizes relevant actions and events in a temporal order to help identify if and how scale-up is on the right track, with an eye towards prospective sustainability. Such sequencing can provide significant leverage and support for project teams and evaluators to enable them to causally trace complex change processes and produce plausible explanations when concrete outcomes are still unknown. The approach also has points of contact with other theories and frameworks that focus on dynamic relationships, such as collective governance (World Bank Group, 2017) and social contract theory for development (Cloutier, 2021), as well as key literature in social science focused on researching and causally explaining complex phenomena (including Pierson, 2004; Guergovich, 2022a).

This paper is targeted primarily for monitoring, evaluation and learning (MEL), and social accountability practitioners interested in or grappling with how to better assess and evidence sustainability and scale-up of specific projects. It also provides a practical means to aggregate that evidence at the program level, in order to demonstrate higher-level impact, enable comparisons, and build testable hypotheses about what works and in which conditions.

The next two sections provide a brief overview of the two key building blocks of this paper.

Section 2 explores the revised OECD DAC evaluation criteria, particularly the sustainability criteria, and the linkages with and relevance to the approach presented in this paper.

Section 3 presents the resonance pathway to scale (Guergovich et al, 2022) and its features.

Sections 4 and 5 present and justify the development of the innovative relational rubric methodology touched on above. It explains how such a method can effectively and practically monitor, evaluate and learn about how scale-up of social accountability programming happens, with a eye towards sustainability.

Section 6 presents findings from applying the relational rubric method to the GPSA sample. The examples and evidence demonstrate that localized projects of three to five years with budgets of less than one million USD can produce processes and outcomes which contribute to many forms of sustainability, including via scale-up, in different settings. This section will likely most interest social accountability practitioners and their funders.

Section 7 concludes with reflections on the promise of a shift of focus to prospective sustainability and the resonance pathway to scale. It proposes ways that the relational rubric approach can help the GPSA and potentially other practitioners and funders better design, monitor and evaluate complex, systemic social accountability projects and portfolios in real-time. It suggests how such learning can build new

knowledge about social accountability sustainability and scale in its many forms, and challenges ill-fitting methodological paradigms and erroneous claims about the limited long-term value of this work. Key recommendations for collectively meeting these aims across the social accountability field are presented.

Annex A presents a deep dive into cross-fertilization across World Bank's [Global Partnership for Social Accountability](#) and beyond. Using this approach, the program delivered on its corporate mandate by producing knowledge that no single project would have produced on its own.



© Dominic Chavez / World Bank. Further permission required for reuse

02

—

Shifting Towards More
Meaningful Monitoring
and Evaluation of
Relational Social
Accountability
Processes within
Dynamic Local
Systems

In 2017, Caroline Heider, then-Director General Evaluation at the World Bank Group argued that the time was ripe for the evaluation community to revisit the evaluation criteria that most development organizations use. She explained that *“development practitioners, as much as evaluators, know that development processes do not follow such linear assumptions. Instead, one action might cause a number of reactions that have effects in rather diverse ways. Hence, we need to develop evaluation models that capture the effects of complexity to inform policymakers and practitioners about the actual effects of choices they make and actions they take”* (IEG, 2017). Two years later the OECD DAC revised its evaluation criteria to assess interventions, taking those insights into account (OECD DAC, 2019). Of particular relevance to this paper and its findings is the revised sustainability criteria, as presented in Box 2.

Box 2: OECD DAC Definitions for Sustainability Evaluation Criteria

Conditions for actual sustainability: This examines *“the extent to which any positive effects generated by the intervention demonstrably continued for key stakeholders, including intended beneficiaries, after the intervention has ended. Evaluators can also examine if and how opportunities to support the continuation of positive effects from the intervention have been identified, anticipated and planned for, as well as any barriers that may have hindered the continuation of positive effects. This can support findings that demonstrate adaptive capacity in an intervention where it was required”*.

Prospective Sustainability (or the future potential for sustainability given factors in the operating environment that could favor sustainability): *“Examining prospective sustainability entails a slightly different approach. An evaluation examining the future potential for sustainability would assess how likely it is that any planned or current positive effects of the intervention will continue, usually assuming that current conditions hold. The evaluation will need to assess the stability and relative permanence of any positive effects realized, and conditions for their continuation, such as institutional sustainability, economic and financial sustainability, environmental sustainability, political sustainability, social sustainability and cultural sustainability”*.

Source: OECD 2021, p73.

The approach for monitoring and evaluating scale-up with an eye towards sustainability, operationalizes two important insights that are embedded within the revised criteria.

Firstly, the new criteria introduces a systemic lens, asking evaluators to consider how interventions fit with the system in which they are implemented and to what effect. A systemic lens focuses on interactions of a wide range of actors in a system, rather than a narrow focus on the siloed actions of the project’s direct civil society implementers. This lens is essential to capture scale-up processes because they often rely on downstream actions taken by others in the system, rather than the ongoing dependence on civil society actors to continue leading this work indefinitely through more interventions. These actors include public sector institutions, funders, and other development agencies that adopt, adapt and/or sustain elements of social accountability processes in different ways after a project ends. This paper illustrates the many ways in which different actors in a complex governance system can support, trigger, and contribute to scale-up in ways that strengthen local public service delivery and their respective systems. Secondly, the updated OECD DAC criteria provides a way out of another problem found within social accountability evaluations and evidence: the ‘absence of evidence versus evidence of absence’ dilemma. This often leads to the erroneous conclusion that if lasting change cannot be evidenced or demonstrated during or right at the end of a project, then movement towards scale-up or other forms of sustainability are not happening and will not continue. When there is high uncertainty and insufficient time has elapsed to observe tangible sustainability, evaluators should consider and use methods that can assess the conditions required for actual sustainability and the likelihood of prospective sustainability, in a wide range of different forms.

The overall methodological framework and relational rubric presented in this paper aligns with and furthers these concepts by integrating and building evidence for the resonance pathway to scale.

03

—

The Resonance Pathway to Scale for Social Accountability

This paper follows and builds on the work of Guerzovich et al (2022), who argue that scale is a complex, relational process. This is often misunderstood in social accountability literature which also fails to acknowledge the important role of social learning and compromise in fostering sustainable results.⁵ To help fill this evidence gap, the GPSA teamed up with World Vision and several experts and colleagues in the social accountability field to research and inform a theory of change focused on scale-up. This theory of change accounted for emergent insights, evidence, and experiences from a wide range of social accountability programs of the GPSA, World Vision and CARE International, amongst other partners.

As presented in Box 3 and Figure 2, several pathways to scale in social accountability were found, including:

- 1. **Replication of best practices**
- 2. **Resistance**
- 3. **Resonance**

This research claims that the resonance pathway has been largely ignored in literature and evaluations to date; this claim is further validated by this paper and the evidence of applying the relational rubric approach. Addressing this theoretical blind spot and evidencing how a resonance pathway works with more fit-for-purpose ways to assess it, will enable social accountability practitioners and evaluators to fill these evidence gaps. It will also help change faulty narratives about the potential and limits of social accountability work.

Box 3: Pathways to Scale for Social Accountability

Guerzovich et al (2022) argue that there are at least three major pathways to scale, based on their research and experience in the field:

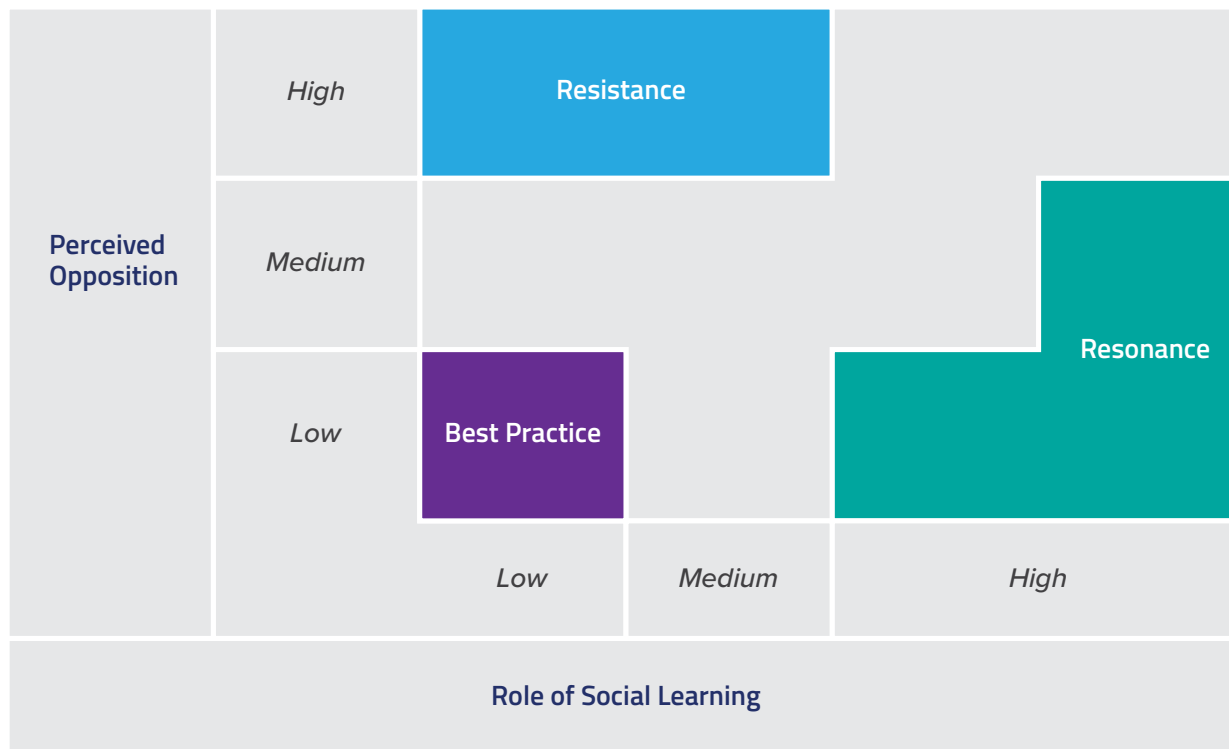
- 1. The **replication of best practices pathway**, whose main anchor is technical expertise and 'rigorous' knowledge.
- 2. The **resistance pathway**, through leveraging the countervailing power of resistance to power and opposition.
- 3. The **resonance pathway** which seeks resonance and best fit with existing public sector efforts.

The first two pathways are commonly assumed in the social accountability literature. However, practitioners' experience often reflects the third one, the resonance pathway. Unlike the others, the main thrust of this pathway is **social learning**. The expectation here is that social accountability work scales-up based on deliberation, compromise, and coordinated collective action among diverse actors. The underpinning logic is that social accountability processes contribute to overcoming the challenges of collective action in a game theoretical sense (Ostrom, 1990; also see World Bank, 2017). That means by enabling a group of individuals to organize and work out how to make the most of a situation (e.g., insights learned by implementing social accountability in select locations), they can create shared gains (e.g., using those insights to inform decisions in other locations) through loose coordination and collaboration.

Each pathway places different emphasis on the **dividends derived from conflict** and on the **promise of social learning** to resolve collective action problems (see Figure 2 below).

5 Social learning for the purposes of this paper and the pathways is a flow – or a chain of events – that involves people engaging with each other and which leads to a change in something they care about. See (Wenger-Trayner, 2014).

Figure 2: The Role of Opposition and Social Learning in Pathways to Scale for Social Accountability



Source: Own elaboration

The resonance pathway and relational rubric method also account for how different actors in a complex governance system can support, trigger and contribute to scale-up in ways that strengthen local public service delivery and their respective local systems. This is a dynamic that existing research and evaluations in the field often miss when they equate social accountability scale with:

- **An overly narrow stakeholder focus**, e.g., only the implementing civil society organizations or community groups.
- **An overly limited range of applicable actions**, e.g., continuation of existing projects, short-term outputs, advocacy, and campaigning.
- **Unrealistic expectations for ambitious results over a short time frame**, e.g., ‘all-or-nothing’ dramatic changes, such as wholesale replication or complete institutionalization of social accountability processes without adaptation, incremental progress, fits and starts, and gray areas (for alternative approaches see Integrity Action, 2020).

Instead, the resonance pathway and the relational rubric recognizes that social accountability scale-up depends on the interactions between and actions of several actors within a given system. This thinking also asserts that there are many discrete elements (or components) comprising a social accountability process. Social accountability interventions are led by civil society, but scale-up often implies negotiation with, adaptation and application of social accountability processes in many forms, by other actors in the system, rather than uptake through replication of the whole process. For example, imagine a collective social

accountability process designed, facilitated and implemented in a few schools through a funded intervention of a civil society group. Rather than expecting this civil society group to fund and continue their work in these same roles indefinitely, the learning and evidence from the intervention can inform how other actors within a school district can implement their own version of such a process in more schools in the future. A systems-thinking lens is required to recognize these nuanced dynamics at play. Due to emerging contextual changes and uncertainty, adaptations to a social accountability process are acknowledged to be common as well as potentially desirable. This supports a narrative for long-term, lasting changes that places local actors as the drivers at the center of their own context-specific development stories.

Taking this view, the goal of sustainable social accountability interventions, scaled or not, should be to contribute towards a stronger local system that is constantly, if imperfectly, *“innovating in terms of how people participate and how those in power are accountable to the society they serve”* (Jacobstein, 2019). Furthermore, the GPSA’s learning and experience over the past decade reinforces that a key ingredient for sustainability gains in complex governance systems is the collective action of civil society, citizens and community groups working with public service actors to jointly organize and solve problems in a way that is suited to their context. Enabling and expecting adaptation to specific local spaces and over time, and the associated ongoing experimentation and learning, are all critical for both delivering and assessing fit-for-context social accountability outcomes and their sustainability.

The GPSA usually tailors its call for proposals to fit World Bank strategies in countries whose governments opted into the program at a given point in time. This may suggest social learning is not at work in the scale-up of social accountability interventions from the ground upwards. However, it is important to emphasize that, ultimately, these are civil society-led processes that operationalize broad parameters and seek to engage local bureaucrats and officials. This is different from the traditional approaches of other World Bank operations that are anchored in the interface between national governments and World Bank staff. Furthermore, the government officials at the central level who opt into the GPSA program or engage in a World Bank operation are not the same actors that engage in a GPSA social accountability project design and implementation at local levels. Even during implementation, key engaged stakeholders often change over the course of the project. Therefore, the uptake of social accountability processes outside of the boundaries of the project is often facilitated by people who were involved in the project at different stages. This is especially relevant in contexts of instability that experience frequent changes within government and civil society, whereby public service officials and civil society members often shift between roles, levels, and organizations. GPSA evidence shows that these actors bring specific learning, capacities, tools and other elements of social accountability processes with them. And that these actors apply adaptations of social accountability processes in their new government posts, civil society or funder organizations and/or require the original designers to consider and accept (or not) compromises as part of the ongoing process.

Box 4: The Role of and Contribution to Local Systems in Social Accountability Thinking and Practice

For the purposes of this paper and its findings, a system is defined as the interconnected set of factors (policies, practices, resource flows, relationships and connections, power dynamics and mental models) that jointly produce a development outcome – the whole is greater than the sum of its parts.

Traditionally, social accountability research and evaluation have focused on standalone interventions, with local systems and their components being of secondary concern. However, more recently, there is a growing number and diversity of schools of thought about ways of thinking and doing social accountability mindful of local systems. According to USAID, a local system refers to actors in a partner country. As these actors jointly produce an outcome, they are 'local' to it. Development outcomes may occur at many levels - local systems can be national, provincial, or community-wide in scope. Using this approach means relying on that local system to produce desired outcomes.

As will be discussed below, there is also evidence about how social accountability contributes towards stronger local systems. From the perspective of international actors, strengthening a local system means building up the capacities of many local actors from government, civil society, communities, and the private sector - and the system as a whole.

Source: Own elaboration

All these interconnected and fluid dynamics also means that scale-up is not guaranteed. The inherent uncertainty involved in complex governance processes and contexts also means that outcomes are likely to vary in form and significance, as the examples in this paper illustrate.⁶ The authors argue that this is why sustainability via a resonance pathway to scale is a legitimate framework for conceptualizing and assessing the outcomes of social accountability projects. And it is also necessary if relatively small projects are to contribute to scale-up in complex dynamic systems. Table 1 synthesizes key insights about the resonance pathway to scale, drawing on the final research paper as well as a series of dissemination blog posts from sector experts (see Guerzovich et al, 2022; Guerzovich, 2021c). These are illustrated by examples and evidence from the GPSA portfolio. As further unpacked in the table, resonance captures the idea of an iterative process of deliberation, compromise, social learning, and collective action through which scale-up happens, with fits and starts (see Guerzovich et al, 2022, and Aston 2022).

6 In technical terms it is possible to assess whether scale up (effect) occurred via resonance of a social accountability project (cause) through theory-informed causal analysis without assuming that endogeneity is a problem or pre-determines results such as isomorphic mimicry (Andrews et al, 2017). Project design seeks to increase the chances that uptake happens but does not determine the results. More generally, on this issue in social science research design see Meadwell, 2022.



© Dominic Chavez / World Bank. Further permission required for reuse

Table 1: Lessons about the Resonance Pathway to Scale from the GPSA Portfolio and Learning

Theme	Lesson	Examples and Evidence from the Field
<p>A focus on the problem that needs solving</p>	<p>The GPSA and other social accountability practitioners’ main aim; focus for sustainability and scale-up; and indicators are centered on the relationships, norms, resource flows and other factors through which actors in the system can contribute to scale and sustainability over time, including as circumstances shift. This approach means that continuity of a particular civil society led tool, project, or even brand that may have been fit for solving a specific problem at a point in time is not expected. Instead, the insights emerging from the collaborative social accountability process inform other actors’ actions. This often entails adaptation and some renegotiation of the original method or tool (e.g., scorecard). This is less of a concern if the adaptation helps to solve the given problem.⁷</p>	<p>In Georgia, the concrete ways through which Save the Children-Georgia, CIVITAS, and partners contributed to implementing the Early and Preschool Education Law at the municipal level provided useful insights that were fed back into the ongoing education policy-making process in country. Insights and relationships from the project were also instrumental in supporting an improved COVID-19 pandemic response when schooling went virtual. However, providing insights for different kinds of problems needs different kinds of responses and listening to different stakeholders – in other words the process of uptake was subject to emergence to develop solutions fit for the job in each case.</p>
<p>A resonance pathway to scale</p>	<p>GPSA projects are likely to contribute to scale-up by promoting social learning, deliberation, compromise, and collective action. The resonance pathway to scale, although new to the social accountability theoretical evidence base, (Haldrup, 2020) reflects the lived experience of many practitioners (including but not limited to GPSA partners in Moldova, Georgia, the Dominican Republic, the Democratic Republic of Congo (DRC), and Mongolia).</p> <p>Collaborative social accountability processes in these cases seem to have enabled groups of individuals to organize and work on making the most of GPSA-funded projects to create shared gains beyond the project time frames and geographical limits.</p>	<p>In Sud Kivu in DRC, insights emerged from activating Village Health Committees through the Cordaid-led GPSA project. These became useful for other donors who used lessons to inform their own programming –they adapted the lessons to their own organizational priorities and circumstances rather than pick up and fund the GSPA project.</p>

⁷ Integrity Action (2021) calls this a process view of sustainability. This is shared by others in the social accountability space.

Theme	Lesson	Examples and Evidence from the Field
<p>Resonance has potential when there is appetite for solving problems with others</p> <p>(Guerzovich et al, 2022, and Guerzovich, 2022b)</p>	<p>Scale-up via resonance seems to be possible in contexts in which there is some mutual appetite for stakeholders to solve problems with others, even if there is initial skepticism.⁸ It seems to be more likely when stakeholders have prior experience engaging in dialogue across the state-society divide and the capacities and trust associated with it.⁹ It is also harder to pivot to resonance when actors have a history of confrontation and mistrust (Aston and Zimmer Santos, 2022).</p>	<p>Rudy Prawiradinata, a Senior Advisor to the Minister of National Development Planning in Indonesia, shared that he was considering how to improve frontline service delivery through citizen engagement, but was concerned about resistance from local authorities as well as raising citizens' expectations without having the capacity to meet them. That is until he talked with stakeholders in communities where Wahana Visi implemented its Citizen Voice and Action project. Then, Prawiradinata realized that there were ways to use insights from this project to achieve his goals, by informing a component of another funding facility (Kompak). The work would eventually be adapted, funded and implemented by the Asia Foundation, with initial support provided from Wahana Visi but no long-term engagement from the organization in implementation (see Annex A and Kompak, 2018).</p>
<p>High levels of perceived opposition (either by government or civil society) need to be overcome to enable resonance</p>	<p>The GPSA identified early on that it would focus its grant-making on targeting concrete problems that actors in specific countries prioritized as fertile ground for joint problem-solving (see GPSA, 2020). In Paraguay, for instance, it focused on addressing shortcomings of the country's conditional cash transfers program (GPSA, 2019), while in Tajikistan it focused on improving community-based monitoring standards for the water and sanitation sector (GPSA, 2018). Tailored context and stakeholder engagement helped to ensure that barriers to collaboration would not be too high nor undermine possibilities for multi-stakeholder social learning, across most of the GPSA's portfolio. It is important to note that there were instances when dialogue broke down but could be renewed (as reflected in the example).</p>	<p>The mid-term and final evaluations of the first GPSA project in the Dominican Republic suggests that pivots from more confrontational to more collaborative approaches are possible over the course of one project. However, behind the scenes this process was burdensome and risky. Organizations predisposed to open confrontation to 'open a door' continued to do so, even when the door was already opened. This meant that time, resources, and opportunities were lost and the risk of losing the trust of officials who opened the door had to be proactively mitigated. There were other cases where civil society groups considered that they were in a zero-sum game with their governments and walked away from the funding and the project.</p>

8 This skepticism is one example of the factors that enable evaluators in specific projects to assume a cause-effect relationship at work in resonance, rather than a situation in which uptake is determined to happen by project design.

9 Organizations that have long-term trajectories in a single site often build these bases in a project cycle and can reap the benefits and raise their ambition in subsequent ones. See (Guerzovich, 2022) on Pact's portfolio.

Theme	Lesson	Examples and Evidence from the Field
<p>Connectors can contribute to or hinder resonance</p>	<p>The World Bank can play multiple roles to support the application or adaptation of elements of GPSA projects, lessons and processes via resonance. For example, facilitating timely access to public sector actors; convening and brokering; informing, advising or funding public sector or other development partner strategies and operations by linking them with insights from GPSA grants (Guerzovich et al, 2020; Guerzovich and Poli, 2020; Green, 2017). Many GPSA grant partners have benefited from the support of World Bank task team leaders (TTLs) and country and sector teams, as they paved the way for sustainability and scale-up.</p>	<p>The 2021-2022 internal assessment of the GPSA’s results framework indicators (which includes a review of several GPSA project evaluations), highlights the importance of the role of the TTL in brokering entry points and linkages between GPSA projects and relevant public sector actors, programs and policies, to enhance the potential for sustainability gains. The evidence also reflected that when TTL project engagement and support is weaker, opportunities for sustainability and scale-up might have been missed or not leveraged to their full potential (Wadeson, 2022). GPSA project evaluations and documentation from projects in Kyrgyzstan, Morocco, Bangladesh or Ghana, suggest that when TTL interest wanes and World Bank teams can no longer enable social learning, the chain to scale-up is more likely to break down (see Mills, 2019; GPSA, 2019a).</p>
<p>Incremental progress despite fits and starts</p>	<p>The kind of systems change associated with social accountability sustainability and scale-up is not linear. Sustaining meaningful achievements over time (and seeding the conditions for them) depends on many actors, their relationships, and interactions as well as other components of the local system. Fits and starts in ‘resonance-style’ uptake due to systemic factors - such as variations in the support from World Bank teams or other local dynamics - are sometimes temporary, rather than permanent.</p> <p>Most project evaluations do not benefit from delays and cannot demonstrate the extent to which any positive effects generated by projects continue after they end, including ongoing influence over policy dialogue. In these cases, it is important that evaluators consider that gradual, complex transformations are more common than often assumed in the social accountability space (Guerzovich, 2022a). To address this timing problem, as recommended by the OECD DAC and discussed in this paper, the evaluative focus should be on a project’s investments in creating conditions for sustainability during its lifetime as well as prospective sustainability after it closes. This can be evidenced through specific signals (e.g., expressed interest, dialogue, established entry points) in the operating environment that could favor sustainability, and therefore the potential for scale-up too.</p>	<p>The World Bank Implementation Completion Report at the close of the TAME project in Mongolia noted that strong local ownership as well as upfront planning and investments in sustainability and scalability diminished risks over time (Meyanathan, 2021). When the independent evaluation of TAME was delayed and carried out months after the project’s closure, it found that the 31 Parent-Teacher Associations established by TAME were still functioning and playing their roles, continuing to find ways to collectively solve problems at school levels. The delayed evaluation therefore uncovered that while the application of lessons seemed to be stalled at project completion, it was rekindled later. This phenomenon has been observed in other GPSA projects, including in Mozambique, and discussed in the 7th GPSA’s Global Partners’ Forum (GPSA, 2021).</p>

Theme	Lesson	Examples and Evidence from the Field
<p>Sustainability and scale that is guided by localization in dynamic systems and focuses on function, not form</p>	<p>Social accountability scale happens in context, mediated by local actors to address local problems in context. What matters is not just whether the specific tools or scale-up processes have the same form – i.e., that they look the same on paper or across contexts. The very process of scale-up with an aim for sustainability requires adaptations to be effective. Therefore, the key is whether such processes function comparably in practice and produce similar effects.</p> <p>These processes of grounding and adapting interventions to local contexts and enabling the ‘continuous pursuit of improvement, innovation in approaches, crowding in by various actors’ are also associated with greater resilience in dynamic contexts. They often bounce back and continue to evolve and produce results, as opposed to attempts or expectations for wholesale replication and formal continuity of interventions that do not fit the context or fail to account for changes in dynamic systems (see Jacobstein, 2019; GPSA 2021a).</p>	<p>An intervention where there are a strong set of pre-existing relationships and joint engagement practices often requires different levels and types of investments for facilitation than for sites where the same types of groups and relationships must be started from scratch.¹⁰ Similarly, a government seeking to scale-up an intervention across a country, whether in-house or by partnering with diverse civil society groups, often needs to consider organizational circumstances that may be overlooked when the intervention is implemented by a single civil society group (or assessed as if the outcome depended on a single civil society actor).</p>

Source:

Resonance is not the only pathway to scale, but it seems to apply to a broad range of social accountability work analyzed by the researchers who defined resonance; this has also since been validated in learning sessions with others working in the field. In cases where the contextual conditions fail to hold, other pathways to scale may be better suited for reaching more geographic locations, sectors and ultimately people (Guerzovich, 2022b). These include the often referred to ‘best practice’ and ‘resistance pathways’ to scale anchored in technical expertise or pressure tactics, respectively (see Box 4 above).

Collectively these insights inform a nested, mid-level theory of change (see Guerzovich et al, 2022) that specifies these three identified pathways (resistance, best practice, and resonance) to social accountability scale-up. These helped the authors to identify where the GPSA experience may fit, as well as the contextual conditions under which each one may be most promising. This theory of change still merits further research to test and refine it. However, the relational rubric approach and assessment exercise discussed in the next sections has supported this effort by building the evidence base further and testing these assumptions.

¹⁰ This layering phenomenon is common across project cycles in the social accountability space (Guerzovich, 2022)

04

—

A Relational Rubric to
Evidence Sustainability
and the Resonance
Pathway to Scale

Through its 2020 updated Theory of Action and Results Framework, the GPSA provided a formal starting point to systematically evidence the resonance pathway in GPSA projects. It included a specific medium to long-term outcome and indicator on uptake, to assess the many forms of sustainability, specially via scale-up (see Table 2 below). As discussed, success for the GPSA is demonstrated when lessons from or elements of collaborative social accountability inform decisions and actions taken by the public sector and other civil society and development actors beyond an individual project, including after the project has ended. This result is often associated with the uptake of selected element(s) of a collaborative social accountability process, rather than complete replication or scale-up of the entire process (wholesale). Such expectations are unrealistic given the scope, budget, and time-frame of GPSA projects (and many social accountability projects in general). It is also important to emphasize that collaborative social accountability processes do not equate to specific tools or capacity development nor are these the only elements that the GPSA seeks for uptake by other actors. While these are important components that can be sustained and scaled based on a GPSA project experience, a social accountability process is much broader, encompassing many elements and examples of ‘what counts’ for sustainability. This section and Box 5 unpack this in further detail.

Table 2: The GPSA’s Results Framework Outcome and Indicator on Uptake (i.e., Sustainability and Scale)

Outcome	<p>Elements of collaborative social accountability processes are taken up by public sector institutions and other relevant actors beyond individual GPSA projects.¹¹</p> <p><i>Other relevant actors can be INGOs/CSOs, World Bank teams, funders.</i></p>
Indicator	<p>Percentage of GPSA grants in which public sector institutions and other relevant actors seek to:</p> <ol style="list-style-type: none"> 1. use substantive lessons for improvements of targeted policies, processes, and mechanisms; 2. apply or sustain elements of collaborative social accountability processes after life of the project; 3. adapt insights from GPSA projects to scale them through programs or policies; or 4. apply elements of collaborative social accountability processes in additional localities or sectors. <p>Note: this can be done through the government’s own reform program, donor-funded programs, or World Bank-financed programs.</p>
Examples	<p>Health sector: Number of priority stakeholders, including local hospitals, public health sector institution officials (central, regional, district and/or village), CSOs, and World Bank team that commit to applying elements of the project’s collaborative social accountability process in additional localities after the project ends (i.e., scale).</p> <p>Education sector: The Ministry of Education uses lessons from the project’s collaborative social accountability process to improve the ongoing education sector policy reform.</p>

Source: Adapted from Wadeson and Guertzovich, 2023

¹¹ Language related to social learning should be integrated into future revisions of this GPSA Results Framework outcome and indicator, the MERL Guide, and this rubric. The GPSA and its MERL framework are always evolving based on learning and evidence.

During the lifetime of a project and at its closure, it is not possible to have certainty about the future in a complex system (or any system). However, it is both possible and desirable to focus on the likelihood of sustainability, and therefore scale-up as one form of it. As per the revised OECD DAC criteria and guidance, related assessments should focus more on signals for prospective sustainability and uptake, such as whether a project has been moving along a sequence of steps that can contribute to scale-up, even after project closure. Due to the sequential nature, monitoring data from the whole project life, and triangulated evidence at the final evaluation stage are both critical. Furthermore, ongoing attention to monitoring data helps project teams to identify, plan for and build opportunities that can support the continuation of positive effects of a project, from the point of its design, while also mitigating barriers and risks along the way.

For example, the mid-term evaluation of TAME project in Mongolia suggests that relevant and effective monitoring, evaluation and learning for action can help projects to more systematically anticipate and integrate specific elements to enhance the prospect for uptake in the future. This intentional practice and foresight can also preempt discontinuity in uptake processes after project closure.

The GPSA's internal monitoring, evaluation, reporting and learning (MERL) system and recently published MERL Guide (Wadeson and Guertzovich, 2023) have built this focus into the indicator and its respective assessment approach. The 'seek to' is a key part of this indicator statement and also operationalizes the OECD DAC's guidance. The uptake of elements of collaborative social accountability processes facilitated through GPSA projects is contingent, in that it can be introduced but not sustained by the project after its closure. The key to evidencing the likelihood is to investigate if concrete steps have been taken by the project, such as: ongoing dialogue with relevant public sector officials and World Bank operations staff over time; identifying potential entry points where elements of the project's collaborative social accountability process can live and grow in future; and the steps stakeholders take to compromise and to leverage those points. This could be through substantive practical forms via another program or reflected in government policy changes or reforms.

Using this relational rubric for measuring the conditions and likelihood prospective sustainability, relationships, associated capacities (especially adaptation) and systemic factors are at the core of evidencing scale-up with an eye towards sustainability in the GPSA model, as in other innovative indicators used in the social accountability field.¹² The project trajectory should ideally demonstrate that this has been an ongoing process from the onset, driving the potential for uptake forward, and course correcting as required, based on solid learning and evidence. Accordingly, the relational rubric is designed to support projects to monitor and evaluate this process and the associated causal sequence of events.

Resonance, as explained above, captures characteristics of scale-up of social accountability (non-linear and emergent, multi-dimensional, the product of multi-directional interactions) entailing processes that often look different in various local systems. They are contingent on localized relationships, social learning processes, and give and take among other factors. This requires a shared understanding about what constitutes progress and success to ensure measurement of what is intended (i.e., construct validity). Therefore, it's critical that evaluative judgments are transparent and clearly understood.

Rubrics, which are a form of qualitative scale to denote levels of performance and support assessment, explain what the standard means and clarify the reasoning behind an assessment. They are a useful instrument to deal with the challenge of assessment of a relational pathway such as resonance. As social accountability evaluator Tom Aston has argued, rubrics *“provide a harness but not a straitjacket for assessing complex change and they help stakeholders build a shared understanding of what success looks like”* (Aston, 2021).

12 For example, Pact monitors the social capital of organizations as key to sustainability and Integrity Action implemented an iterative approach to monitor sustainability. For other approaches see Guertzovich, 2022.

Figure 3 introduces the five-point relational rubric¹³ that the authors developed to transparently assess the many possible outcomes for sustainability, including via scale-up, in line with the resonance pathway and the GPSA's Theory of Action.¹⁴ It was tested across a sample of 15 projects supported by the GPSA. It will be used for current and future GPSA projects and to aggregate and compare results at the portfolio level across time. It is important to note that this rubric can be applied at different stages of a project, not just at the final evaluation, for example, to monitor prospective sustainability throughout the project or for ex-post reviews, by slightly adapting the language in the rubric to clarify the timing.¹⁵

13 To learn more about the rubric and guidance for its application during both monitoring and evaluation phases, see Wadeson and Guertzovich, 2023.

14 This rubric was specifically designed for assessing the evaluation criteria of sustainability as conceived by this paper, not for other evaluation criteria (e.g., OECD DAC criteria of effectiveness, impact). If additional criteria are part of a MEL system or specific evaluation, then other suitable evaluative tools and methods would be needed to complement the rubric.

15 For example, a GPSA internal learning exercise in 2022 collected long-term results from a sample of closed GPSA projects (unpublished). The sample included 14 of the 15 projects used for the initial rubric test and findings presented in this paper.

Figure 3: Sustainability Relational Rubric Levels with Criteria



Source: Adapted from Wadeson and Guertzovich, 2023



© Simone D. McCourtie / World Bank. Further permission required for reuse

05

—

Reflections and Guidance on Using This Relational Rubric

While presenting empirical findings of results is always a key priority and interest, the process that enables evidencing them is also important to explain, because the ‘what is the how’. Previous theoretical and methodological choices have created blind spots that pre-empted useful and fit-for-purpose evidence for sustainability and scale-up of social accountability work. This section provides a range of partly overlapping considerations associated with the definition and application of this rubric to address challenges and misconceptions and to provide more guidance.



A politically minded, relational rubric: Throughout the life of GPSA projects, grant partners seek to create buy-in and demonstrate the value of collaborative social accountability processes to a range of relevant actors who can support or directly ensure downstream uptake which often calls for repeat interactions and embracing compromises. In doing so, these actors often develop a stake and ownership in the process, becoming more capable and likely to promote collaborative social accountability. As a result, they may actively seek opportunities for scale whether through new or existing policies or programs or to apply insights and lessons in new localities and sectors. These actors also play insider influencing roles, inaccessible to many CSOs due to common power asymmetries. As such, they can better identify where uptake and partial adoption or adaptation of collaborative social accountability processes are possible, and effectively support scale-up for long-term sustainability. Achieving this requires strong political acumen, relationship building skills, and access on the part of grant partners.

If relevant actors are invested, they may choose to put their own ‘stamp’ on a collaborative social accountability process or adapt it with an iterated model when scaling, so that it resonates with their own perspective and place in the system. This product of deliberation, compromise and loosely coordinated action is considered a success in the GPSA’s Theory of Action. Such a case would be a strong example of impact, even if the project’s specific contribution is less visible, and the form of the process looks quite different than what it was during the project.



Identify logical sequencing and incremental steps: The expectation and existing evidence points to the tendency of scale-up to be incremental. A critical element of enabling this is the project’s efforts to engage and find avenues for sustainability should take place throughout the project life, not just at the end. GPSA learning over time has shown that there is often a certain sequence of events associated with uptake for scale-up. There is a long tradition in comparative institutional and political analysis to study ‘cases’ that are decomposed into a sequence of events and whose *“causal claims rest upon the inferences derived from the analysis and comparison of those sequences”* (Falleti and Mahoney, 2015). This ‘comparative sequential’ method, which *“can and must encompass more specific methods of cross-case analysis and within-case analysis,”* informed the development of the relational rubric.¹⁶

The rubric established a logical sequence of events to be pro-actively driven by a project to enhance the likelihood of sustainability, including scale-up. Its performance levels are directly related to each stage in this sequence. This starts from no evidence of any interest expressed, dialogue with or actions taken by relevant priority stakeholders regarding sustainability and scale of the social accountability process. It then moves through a chain of increasing levels of interest, identification, dialogue, and actions, requiring sufficient supporting evidence or insights for substantiation. (of element(s) or insights from a collaborative social accountability process by any priority stakeholders or institutions). It sets out concrete, observable steps that could be causally linked to a project’s efforts, evidencing its contribution.

In this way, the relational rubric is designed to help project teams and evaluators to effectively trace and evidence the sequential steps that often play out in a causal manner. This specificity and ongoing evidence collection also supports project teams to make informed course-corrections when projects are not on track.

¹⁶ Process tracing informed our approach to theory building and testing theory within cases, while a transferability lens, grounded in simple matching tools and other methodologies, was used for meaningful aggregation and comparison of complex portfolios. On the former, see Section 4; on the latter see Wadeson et al, 2020.



Define reasonable limits for the rubric scores and criteria: The 1-5 scale in the rubric provides criteria to guide scoring of the indicator with a sequential logic, where each step is assumed to be a steppingstone to the next. Each numerical score corresponds to a percentage, with all percentages above zero representing different degrees of partial uptake. However, full uptake is not included as this is unrealistic. Setting this limit ensures a realistic ambition for projects and their assessments. By design, the GPSA does not expect that individual and relatively small-scale projects will lead to sustainable collaborative social accountability processes in the form of complete adoption, continuation, replication and/or extensive scale-up.



Ensure realistic expectations for scores and expect uncertain, emergent effects: A score of 2 or 3 reflects a positive outcome. A score of 4 or 5 would be regarded as a significant success, but also one that is quite challenging to achieve within the time-frame and scope of most GPSA projects – which makes it important to judge the direction of travel in context. Expectations should be tempered given the timescale and resources of a given grant as well as the nature of progress – often incremental. For example, it could be reasonable to celebrate a score of 2 earlier in a project but to expect a 3 or 4 by its end.

Scale-up is an experimental and highly contingent result, it requires difficult changes to result from the interactions of multiple stakeholders and momentum to be maintained (although often with stops and starts) over time-frames that are usually longer than a project cycle. Focusing on ‘ideal’ results can obscure learning about plausible, partial results, which had been identified by the GPSA community as most relevant to learning and course-correction. The relational rubric is designed to guide monitoring and evaluation practice in a more realistic and fit-for-purpose way considering partial uptake of specific elements with modifications as ‘success’ to be evidenced, learned from, and shared. It is therefore important to cast a wide net on different potential outcomes for sustainability in a range of different forms as success is manifested in diverse ways. Box 5 provides examples of ‘what counts’ and what should not be considered as evidence of scale-up, based on GPSA learning and evidence, and the findings of this exercise.



Focus the relational rubric explicitly on processes and functions, not tools (see Wadson, 2020): As previously discussed, scale-up happens in context. It is mediated by local actors who seek to address local problems suited to their given locality and sector. The process brings together unique combinations of stakeholders, dynamics, norms, perspectives, and experiences, amongst other variables. Therefore, these processes will vary widely in their form between contexts to be locally relevant. As a result, the specific form the exchange takes or the design of specific components of the process is far less important than the way the process is meaningfully adapted to the context to support deliberation, compromises and, eventually, coordinated action.

For example, an evaluation of GPSA projects in Malawi explains that when teachers who participated in social accountability processes were transferred to new schools, they *“inspired by the project activities became harbingers of the social accountability initiatives ... at their new schools. Similarly, the [primary education advisors] in targeted zones had participated in capacity building initiatives on social accountability principles and practices and observed them at work in the targeted schools. They took the messages to non-project schools within their zones through their advisory roles”* (Chingaipe et al, 2022).¹⁷

In different Malawian schools, the transferred teachers and primary education advisors look different, as do their specific activities, but they are playing the same role of supporting uptake of lessons beyond the initial sites of project implementation. This result seems to have been obtained

17 This evaluation was concluded after the initial testing of the rubric, but further validates the findings of this note.

in other GPSA projects not included in this initial rubric testing.¹⁸ In other projects, government authorities or World Bank staff performed this function and paved the way to scale, as will be discussed more in the next section.

Focusing on function over form starts by defining standard but broad concepts, such as uptake, sustainability and scale-up with an eye towards sustainability. In many projects, GPSA stakeholders defined both old and new concepts together, considering emergent practices, evaluations as well as research in the sector. Their aim was to craft clear, explicit definitions to support common understanding about what exactly the GPSA intends to measure and learn about collectively, before determining the 'how'. Such definitions are also important to support the transfer of key ideas to different project contexts more easily, and to ensure similar dynamics are being assessed (i.e., construct validity). This isn't meant to prescribe, but rather to ensure consistency while consistently drawing on multi-directional learning and dialogue with partners on the ground and evaluators (such as discussed in Annex A). While there are no perfect definitions and concepts that can evolve with learning and practice, striving for 'good enough' is reasonable, while also being careful about conceptual stretching.¹⁹



Functional equivalent indicators are key: After concepts are defined for common understanding, establishing a set of core indicators to operationalize these concepts is important, with specific guidance on what is (and is not) essential to document, monitor and evaluate across projects. This enables aggregation and comparison at the program level for richer more systematic evidence and learning over time. These indicators should be linked to the theory of action or theory of change of a given program, articulating how and when change is expected to happen on the pathway to scale.

While these indicators will be localized to each project, they should be standardized in a few ways. This includes using the same units of measurement and representing the same core concepts and assumptions for how change is expected to happen over time. These are referred to as 'functional equivalents' - akin to the expression 'comparing apples to apples'. Ensuring functional equivalents enables reliable comparison across projects in different geographies and sectors and also allows for aggregation at the program level (see Wadeson and Guerzovich, 2023). The expectation is that with more comparable data, the appropriate conditions and time horizons for impact can be identified in realistic terms, based on a broad range of examples of 'what counts' for success. For the GPSA, this means that key concepts and elements (including practices, approaches, tools, and mechanisms) involved in collaborative social accountability processes are consistently defined and represented by the role (function) they are expected to play rather than by their exact form. The forms they take across GPSA projects may look different in practice, as they should since they need to be adapted and localized to the project context (Wadeson and Guerzovich, 2023). It is important to note that the process of clarifying concepts and ensuring functional equivalents often requires dedicated MEL staff and external evaluators to work directly with project teams to build this understanding and practical capacity. The level of effort depends on how much project teams will be directly involved in the MEL of a project (the degree of participation).

¹⁸ Tom Aston shared evidence to solidify this result in his independent evaluation of a GPSA project in the Dominican Republic.

¹⁹ Mahon (1993, 845) defines conceptual stretching as "the distortion that occurs when a concept does not fit the new cases."



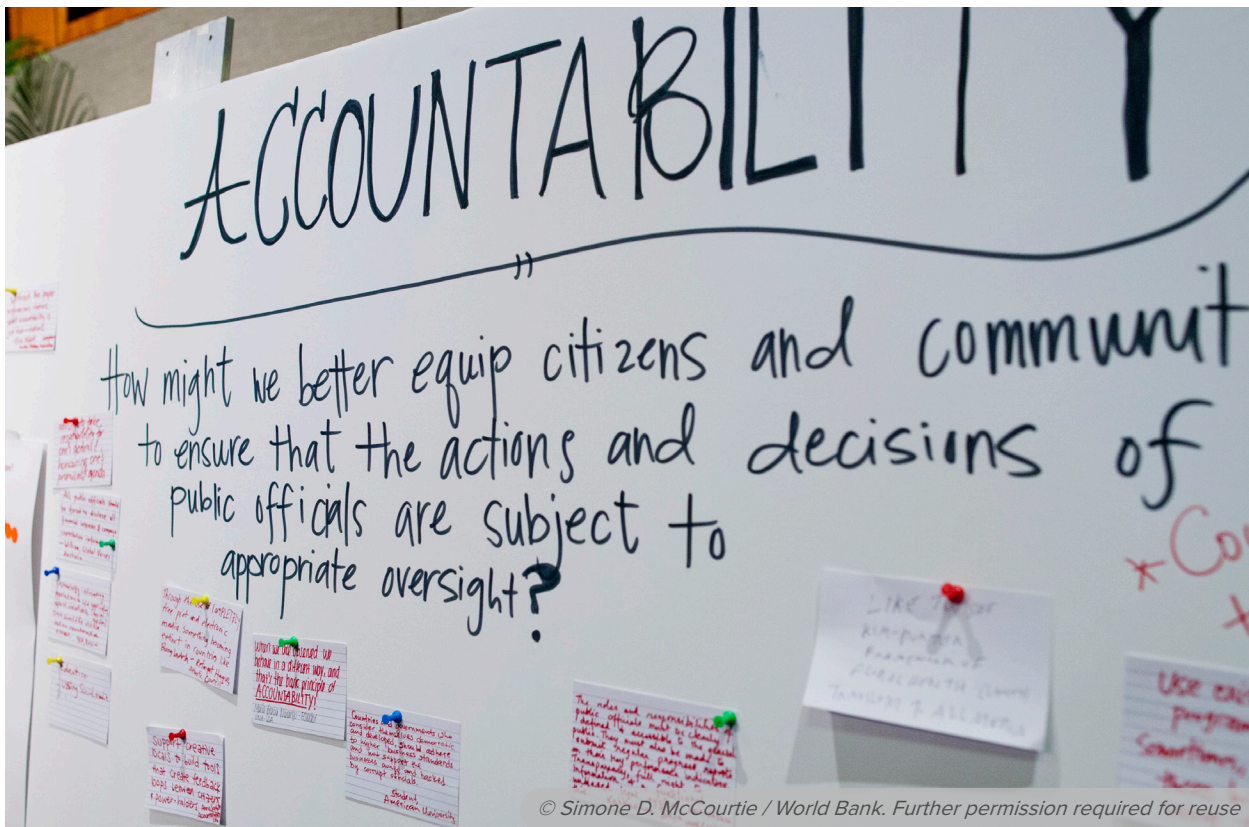
Make the assessment based on triangulated evidence: Ideally this will use monitoring data from project teams as well as their own primary data to make a balanced and robust assessment on the rubric scale. While the rubric application used GPSA projects as the sample and evidence base, it is relevant beyond GPSA projects. Other MEL practitioners in the social accountability field are encouraged to help test, learn from and improve it over time. It is recommended that MEL practitioners who wish to apply it as intended, understand both rubrics as an assessment tool, and the conceptual framework for social accountability sustainability and scale-up that are presented in this paper.



Ensure sufficient evidence quality and range: In addition to casting a wide net for what counts, using the rubric to reach specific scores relies both on the quality of evidence generated by a monitoring and evaluation system and the range of identified examples analyzed in context. A lack of triangulated verifiable evidence of examples with insufficient details on why and how the uptake happened would not contribute to learning about what works for sustainability and scale-up of collaborative social accountability processes.



The use of percentages: Recognizing appropriate expectations for sustainability, a rubric score of 5 or 100% does not equate to wholesale uptake or scale-up. Translating each level in the rubric to a corresponding percentage score provides a common reference point that is easily understood and comparable. Each level in the scale moves sequentially from no uptake to partial uptake, to full uptake (while noting that full uptake does not mean total replication for the GPSA, as discussed). However, the use of percentages is not essential in the use or adaptation of this rubric by others. The criteria for each level of the scale can be adjusted, but the authors advise that the core features of the rubric approach and its conceptual underpinnings are maintained.



Box 5: What ‘Counts’ for Results of Uptake?

Cast the net wide to mitigate blind spots and identify hidden successes:

The rubric embeds the same ‘detective’ approach that allowed the GPSA to uncover tacit knowledge about resonance to date, including actions taken as a result of lessons from collaborative social accountability projects, even in cases where the decision is to not pursue the recommendations as they were written. This means ensuring that a range of results ‘count’ positively when collected and analyzing data. Examples range vastly and can include:

1. GPSA/World Bank project personnel, and representatives from civil society and community organizations and networks integrated into another public sector project or program.
2. Public sector counterparts used lessons to inform public sector reforms and policies.
3. Emulation by local public sector or service providers (e.g., education officials and schools) that observed, adopted or adapted the collaborative social accountability process from a project.
4. The World Bank or other funders used lessons and approaches to advise public sector or other development partners’ programs.
5. The World Bank or other funders financed an adaptation of the project in the same or other sectors.
6. Any observed or reported uptake, sustainability and/or scale-up led by other international non-government organizations (INGOs) or civil society organizations (CSOs).
7. The project actions and trajectory demonstrated ongoing dialogue with key actors (relevant public sector officials and World Bank operations staff) to move the process for potential uptake of collaborative social accountability processes forward.

Be clear about the boundaries of what does not count:

It is important to be clear about events and actions which are not applicable examples of uptake, even though they are sometimes claimed as such. These include:

1. Grant partners share a report or knowledge product and invites key stakeholders to events on learning (an output).
2. Grant partners meet with government, without information about follow-up or subsequent actions taken.
3. Grant partners run a campaign, issue documents and messaging for awareness, advocacy, etc.
4. The media disseminates the content of civil society demands for collaborative social accountability or related advocacy messages.

While these might be useful activities that are part of the project implementation or results in terms of information sharing and dissemination, they do not count as a positive instances of uptake action by decision makers to support sustainability and scale-up. It is the actions taken and/or the use of information shared which is critical for uptake and to enable sustainability and scale-up, as per the GPSA’s conception.

Source: Own elaboration

06

—

Testing the Relational
Rubric to Better
Understand and Learn
about Collaborative
Social Accountability
Scale-up

The relational rubric was tested, using sample GPSA projects, to uncover whether it enables better understanding about whether and how the projects have scaled-up, and whether it validates the GPSA's Theory of Action and the resonance pathway to scale. The test methodology is shared in Box 6.

Box 6: Testing the Relational Rubric on a GPSA Sample

The first step towards testing the relational rubric was to define the criteria for a purposive sample. There were 30 eligible projects within the GPSA portfolio that could be considered. The sampling and analysis were done by a professional evaluation consultant working with the GPSA, who had not had involvement in any of these projects. The scope of the sample also considered the limitations on resources available for this initial test. The final criteria were:

- Completed projects, as these would be the most likely to demonstrate potential for scale-up.
- Projects with several high-quality secondary documentations covering the whole project life.
- Projects with external midterm and/or final evaluations and/or World Bank Implementation Completion Reports (ICR), because these sources often have the most detailed information and include independent assessments, which helps for triangulation and mitigation of bias.

The final purposive sample of 15 projects (50 percent of the sample frame) is presented in Figure 6.

After sample selection, the application of the rubric was piloted in one project in the sample – the TWISA project in Tajikistan led by Oxfam. The following examples of scale-up were identified:

1. Another CSO adopted the project's collaborative social accountability model in other country locations, which were not part of the original TWISA project.
2. WHO Tajikistan used the project's Service Performance Indicators in their project on water supply and sanitation services assessments and water safety plans.
3. The Swiss Development Cooperation Agency supported use of the project's collaborative social accountability model by government implementing agencies that they are funding.
4. The European Commission also supported use of the project's collaborative social accountability model in its other funded projects.

Evidence that supported this was sourced in the project documents:

- The Implementation and Completion Report provided specific details and actors demonstrating actions for collaborative social accountability process uptake or expressed interest/support for it.
- The independent final evaluation found that the project actively created synergies with other programs and actors that could help with uptake (scale-up/scale-out) and sustainability i.e., Tajikistan Water Supply and Sanitation Network, and engaged the Ombudsman presidential appointee.
- The project identified suitable policy entry points to advocate to the government for collaborative social accountability support i.e., the new Action Plan for Water reform signed-off by the highest national authority and the Ministry of Energy and Water.
- The project pursued avenues for long-term sustainability from the onset and throughout, not just at the end.

This combined and triangulated evidence resulted in a score of 5, considered as 100% uptake as per the rubric criteria. The rubric was then applied to the other 14 projects in the sample. The results were aggregated.

Source: Own elaboration

Figure 6: Final Purposive Sample of Projects



 <p>Uganda</p> <p>PROJECT: Uganda Enhancing Accountability & Performance of Social Service Contracts</p> <p>LEAD GRANT PARTNER: Africa Freedom of Information Centre</p>	 <p>Mozambique</p> <p>PROJECT: Social Accountability Knowledge, Skills, Action and Networking</p> <p>LEAD GRANT PARTNER: United Purpose (previously Concern Universal)</p>	 <p>Paraguay</p> <p>PROJECT: TEKOPORÁ - Beneficiaries Empowered to Ensure Social Accountability</p> <p>LEAD GRANT PARTNER: Fundación Comunitaria Centro de Información y Recursos para el Desarrollo (CIRD)</p>	 <p>Ghana</p> <p>PROJECT: Making the budget work for Ghana</p> <p>LEAD GRANT PARTNER: SEND</p>
 <p>Morocco</p> <p>PROJECT: Linking Education and Accountability for Development (LEAD)</p> <p>LEAD GRANT PARTNER: CARE</p>	 <p>Moldova</p> <p>PROJECT: Empowered Citizens Enhancing Accountability of Education Reform and Quality</p> <p>LEAD GRANT PARTNER: Expert Group</p>	 <p>Madagascar</p> <p>PROJECT: Citizen Involvement in Municipal Service Improvement (CIMS)</p> <p>LEAD GRANT PARTNER: SAHA</p>	 <p>Mongolia</p> <p>PROJECT: Transparency and Accountability in Mongolian Education (TAME)</p> <p>LEAD GRANT PARTNER: Globe International Center</p>
 <p>Philippines</p> <p>PROJECT: Guarding the Integrity of the Conditional Cash Transfer</p> <p>LEAD GRANT PARTNER: Concerned Citizens of Abra for Good Governance (CCAG)</p>	 <p>Dominican Republic</p> <p>PROJECT: Prácticas de Buen Gobierno en la República Dominicana</p> <p>LEAD GRANT PARTNER: Oxfam</p>	 <p>Georgia</p> <p>PROJECT: Improving Pre-school Education in Georgia through Social Accountability Processes Project</p> <p>LEAD GRANT PARTNER: Save the Children</p>	 <p>Indonesia</p> <p>PROJECT: Citizen Voice and Action for Government Accountability and Improved Services</p> <p>LEAD GRANT PARTNER: Wahani Visi</p>
 <p>DRC</p> <p>PROJECT: Reinforcing social accountability of health services by supporting health committees and the community diagnosis in Bas Congo and South Kivu</p> <p>LEAD GRANT PARTNER: Cordaid</p>	 <p>Tajikistan</p> <p>PROJECT: Improving social accountability in the water sector (TWISA)</p> <p>LEAD GRANT PARTNER: Oxfam</p>	 <p>Bangladesh</p> <p>PROJECT: Social Engagement for Budgetary Accountability (SEBA)</p> <p>LEAD GRANT PARTNER: Manusher Jonno Foundation (MJF)</p>	

Source: Based on World Bank data

Each of the 15 projects in the sample was deemed to achieve at least partial uptake, meeting a rubric score of 2 (25% uptake) or more. Several examples were identified across the source documents (and in many cases, the final independent evaluations) and triangulated as much as possible with the data available.²⁰

Table 3 shows the breakdown of results by rubric level.

Table 3: Results of the Relational Rubric Testing Exercise on a Sample of 15 GPSA Projects in 2021

Relational Rubric Score	# of Projects in the Sample	Relational Rubric Criteria
1 or 0% UPTAKE	0 projects	No evidence of any use/application/adaptation of element(s) of or insights from a collaborative social accountability process by any priority stakeholders and/or public sector institutions. No evidence of stakeholder interest, dialogue of alignment evidenced.
2 or 25% UPTAKE	2 projects	Evidence of interest by priority stakeholders and/or public sector institutions expressed publicly or privately about learning from a collaborative social accountability process in the project.
3 or 50% UPTAKE	4 projects	Evidence that priority stakeholders and/or public sector institutions have expressed where to adopt, adapt and/or sustain elements or insights from a collaborative social accountability process and how this could be incorporated in some way into other operations, programs, policies (i.e., concrete entry points have been identified).
4 or 75% UPTAKE	4 projects	Evidence of dialogue with priority stakeholders and/or public sector institutions on how to adopt, adapt and/or sustain elements of the collaborative social accountability process in future operations, policies, or programs.
5 or 100% UPTAKE	5 projects	Evidence of actions taken by priority stakeholders and/or public sector institutions to adopt, adapt and/or sustain elements of a collaborative social accountability process in other operations, policies, or programs. Triangulation of data with at least 2 sources of evidence to confirm is required.

Source: Based on World Bank data

²⁰ The 15 projects not included in the sample did not meet the criteria for evidence availability. Therefore, their potential sustainability is unknown; primary data collection and targeted analysis would be required to determine this.

The exercise was limited as it did not allow for primary data collection due to time and resources available, but it provided good enough evidence for the purposes of testing the relational rubric. It was conducted by an independent consultant who was not part of the projects or the GPSA team which helped to mitigate bias. The relational rubric was also reviewed and tested by other evaluation experts to support inter-rater reliability. Testing should be continued as the rubric is used by the GPSA and others.

The testing exercise helped to refine the relational rubric further, and surfaced the following lessons:

- ▶ **It can be applied using secondary evidence. Although evaluations did not explicitly apply the rubric, the retrofitting of data to apply to the rubric was still possible and useful.**
- ▶ **It made sense, helping to build confidence in it as a fit-for-purpose method to assess sustainability and scale outcomes.**
- ▶ **It provided a transferable metric to compare and aggregate these outcomes and the success of different GPSA projects.**
- ▶ **Its logic and format were easy to communicate to others (as was done at the GPSA's 8th Annual Partners Forum, in the Scaling Social Accountability panel session (GPSA, 2022)).**
- ▶ **It validated the assumption about casting a wide net for what counts for success, and captured scale-up in the various forms that a resonance pathway can take.**
- ▶ **It supported the GPSA's inductive-deductive approach to theory building.**

The rubric with key evidence and lessons from using it has been presented at internal and external forums with positive reception from World Bank stakeholders, GPSA grant partners and evaluators outside of the GPSA. While there is clear promise in the process and results, the use of the relational rubric is experimental; it is still being tested. It has been and should continue to be updated with new evidence and learning, including beyond the GPSA.

The wide range of potential outcomes for sustainability criteria to be met means that scores can be aggregated and compared. However, what this looks like in practice will still differ vastly across projects. So, it is important to emphasize that when projects received the same rubric score, scale still looked very different. For example, four projects were given a rubric score of 3 or 50%. This means that there was *“evidence that priority stakeholders and/or public sector institutions have expressed where to adopt, adapt and/or sustain elements or insights from a collaborative social accountability process and how this could be incorporated in some way into other operations, programs, policies (i.e., concrete entry points have been identified).”* To meet this criterion, a wide range of examples were included from these four projects, such as:

- ▶ **Expressions of interest from government/CSOs to use tools and guidance developed by the project (SEND Ghana).**
- ▶ **Likelihood of a governance body mechanism (like a steering committee) to be used in future projects (SEND Ghana).**
- ▶ **Government actors and parliament making regular requests to the lead CSO for inputs on relevant sector matters post-project (SEND Ghana).**
- ▶ **Key international aid actors (GIZ, UNICEF, USAID, IRC) report use of approaches derived from the GPSA/CODESA experience, often after having witnessed it in the field (Cordaid DRC).**

- ▶ **Project managers who exited the project ended up in influential positions in other institutions and, when interviewed, suggested that they were using CODESA ideas in the design of large UNICEF, World Bank, and USAID programs (Cordaid DRC).**
- ▶ **Evidence that GPSA/CODESA health sector experience influenced the set-up of social accountability mechanisms in the education sector (Cordaid DRC).**
- ▶ **Municipalities have pledged to continue the collaborative social accountability mechanism of benchmarking beyond the project. Additional municipalities have expressed the will to participate in the established governance mechanism (Save the Children Georgia).**
- ▶ **Social audit practices are being considered for scale-up to other districts, including nationwide, and to other programs of the Ministry of Social Development (CIRD Paraguay).**
- ▶ **A World Bank Country Partnership Framework incorporated new engagements expected in social protection system effectiveness including social accountability mechanisms based on the project experience (CIRD Paraguay).**

The findings of the test exercise suggest that it is plausible to evidence the realistic contribution to scale-up with an eye towards sustainability of collaborative social accountability processes in projects with a duration of three to five years and budget of less than USD one million, provided projects are assessed with relevant conceptual frameworks and methodologies under qualified MEL staff and evaluators. Those methodologies need to be suitable for exploring and capturing actions taken by priority stakeholders and/or public sector institutions (to adopt, adapt and/or sustain elements of a collaborative social accountability process in other operations, policies, or programs) both within. This can include signals or commitments made by them during the project life for future scale-up. For example, in Madagascar, most participating municipalities (32 of 46) budgeted for the continued operation of the collaborative social accountability processes post-project and the expansion of the approach by new projects and areas by other development actors.²¹

The findings validated the assumptions of the authors and other partners that there are many potential outcomes and forms of scale-up (see Box 5). They also reinforced the need to look beyond numbers and percentages in the rubric when analyzing the data; a relational and systems lens is critical for understanding the ways in which the assets and learning of various local actors improve on approaches in ways beyond those envisioned during the project design. The potential for growth and resilience exists within these emergent systemic dynamics, through continuity and ongoing adaptation of social accountability processes suited to evolving local contexts. Transformation, thus, is not contingent on wholesale adoption of solutions advocated by the organizations that initially designed or piloted the solution. Adaptation, give and take, and social learning are essential elements that enable resonance with others holding different perspectives and ideas to be taken forward. It is important to review the qualitative details and narrative trajectory of each case to understand the nuances and vast range of how sustainability can take shape on the resonance pathway to scale.

An important caveat is that the longer-term trajectory for scale-up, signaled at the end of a project and often reflected in its final evaluation, is not a guarantee. Some processes for scale-up will stop and stall during and after project implementation. Others may resume after years of stalling. Systematically analyzing this variation can offer important insights for how to monitor and evaluate the scale-up of social accountability processes across different contexts.

21 See [Jespersen, 2022](#), explaining that some improvements piloted by TAME have informed the Basic Education Support Project (PASEB II) – co-financed by the World Bank and the GPSA. Also see [Lekweiry and Falisse, 2022](#).

07

—

Concluding
Insights and
Recommendations
for Measuring
Complex Change in
Social Accountability
Work and Beyond

Many of the assumptions at the core of the approach and findings discussed in this paper are consistent with a growing evidence base from broader literature on systems aware work. Commonly used evaluation frameworks are not compatible with the complexity and uncertainty associated with these processes and equate mixed results to lack of effectiveness, contributing to problematic 'gloom and doom' narratives in the social accountability field (Aston, 2021). Efforts to develop and apply relevant monitoring and evaluation approaches have often been seen as too complicated and dismissed or poorly prioritized. However, the emergence of a vibrant community of monitoring and evaluation professionals committed to developing methods that are well-suited to support the complex, systemic and pressing development problems of this era – from the climate crisis to governance failures is changing that trend.

This paper introduced the relational rubric as part of a systematic and operational approach for monitoring and evaluating complex change, applied to collaborative social accountability programming. The rubric is consistent with the revised OECD DAC evaluation criteria for sustainability. The findings discussed show that systematically and causally tracking complex scale-up with an eye towards sustainability is achievable and not as difficult as previously imagined. While using primary evidence in real-time is highly recommended, the authors found that this can still be done in a meaningful way even using secondary data, long after project closure.

The Value of Applying a Relational Rubric Grounded in a Resonance Pathway to Scale

Building upon previous evaluation-action work on pathways to social accountability scale with an eye towards sustainability, this paper focuses on a new relational pathway - resonance. The resonance pathway was previously missed and seems to apply to a broad set of (but not all) social accountability interventions and contextual circumstances. The relational rubric helped operationalize the resonance pathway further, capturing the many forms of sustainability, involving scale-up through deliberation, compromise and coordination of diverse local actors working in complex governance systems.

As hypothesized, these emergent, multi-stakeholder processes of scale up with an eye towards sustainability can be as complex as the systems that they help to strengthen. Yet the rubric demonstrates that they can also be knowable and traceable with fit-for-purpose concepts and tools. It provides a way of uncovering them in a systematic way across very different projects and contexts; at both the project and portfolio-level. Overall, the findings of this inductive-deductive exercise validated the promise of the resonance pathway to scale as well as the relational rubric method for evidencing it – both of which are fit for evidencing incremental, transitory, piecemeal, and intermediary processes, contexts, and outcomes (Guerzovich, 2022a). It demonstrated that projects did move towards scale with a view towards sustainability, along a pathway that is coherent with and strengthens local dynamic systems.

The testing of this relational rubric also revealed that it is possible to find evidence and concrete examples of prospective sustainability in many forms by the time projects end. However, expectations for achievement and/or evidence of sustainability and scale-up during the life of most projects or directly after they end, are both overly ambitious and unhelpful for projects with short time frame and limited budgets (such as GPSA projects with average three-year durations). Therefore, the sequential logic embedded into the relational rubric approach focuses on and provides a practical tool to causally assess the conditions required for actual sustainability and the likelihood (rather than the certainty) of prospective sustainability.

Potential Uses of the Relational Rubric at the Project-Level

The sample data was coded retroactively using a range of independent and internal project documentation. However, the value and utility of the relational rubric can be increased if applied during a project, with real-time targeted MEL to provide actionable information to support teams designing and implementing projects:

- By helping them to reflect on whether they are building a viable approach to achieving (realistic) outcomes for sustainability and scale-up into their project implementation and strategies.
- To spot emergent windows of opportunity, analyze the trade-offs and compromises in different scenarios, and course-correct accordingly. For example, at the mid-point of a project, it is helpful to intentionally reflect on whether potential entry points and relationships with actors in the system who could support uptake have been identified and acted upon, and whether others are emerging following contextual changes. This can catalyze planning and actions by project teams to leverage during the remainder of the project, rather than considering this at end, when it is often too late, and teams are focused on completing implementation and project close-out.²²
- Evidencing challenges go beyond the social accountability field and this framework can also be applied to projects that support other types of systems strengthening. For example, the World Bank supports political economy analysis and other analytical products in several contexts, however it can be quite difficult to measure the uptake of the resulting information that is provided to counterpart operational teams and partners. The relational rubric could potentially help.²³

Potential Uses of the Relational Rubric at the Portfolio Level That Can Support the Wider Social Accountability and MEL Fields

Beyond assessing individual projects, conceptualizing and evidencing social accountability scale-up along the resonance pathway with the rubric method, provides important insights about transferring the results of individual projects within portfolios. A portfolio-level analysis of interventions can produce insights greater than the sum of their parts. The interaction and iteration of multiple evaluations form the basis for theoretical, methodological and empirical innovations that no single project could deliver alone. In turn, this can support the wider social accountability field by:

- Helping both social accountability and MEL practitioners better understand and address the siloes created by project MEL that is unconnected to the wider evidence base, preventing knowledge transfer. The fragmentation found between locally led processes can be overcome through cross-fertilization of results and learning across projects. This can help build stronger theories of action and change and social accountability narratives, especially about scale-up and sustainability.
- Fostering more meaningful comparisons and aggregation of results within an overall social accountability program, such as the GPSA.
- Contributing to improved learning about how sustainability and scale-up happens in the broader social accountability field as interventions induce new interactions, innovation, and changes in local systems. In this sense, the relational rubric provides a useful building block to help address the 'absence of evidence' dilemma found in current erroneous assumptions and incomplete evidence related to the scale-up of social accountability work for long-term prospective sustainability.

22 For detailed examples and guidance on using the rubric for monitoring, see [Wadeson and Guertzovich, 2023](#).

23 Thanks to Mathieu Cloutier for this insight into the promise of the rubric method and it could be useful beyond social accountability work, as per this example about the political economy analysis of the World Bank.



Recommendations

More testing and iteration of the relational rubric, alongside intentional design of projects and evaluations to apply it in real-time with primary data, will lead to a growing evidence base for the resonance pathway to scale, increase knowledge about social accountability scale-up, and highlight the many legitimate forms of sustainability outcomes. The authors propose the following recommendations to enable this in practice:

Investment in targeted and robust research and evaluations to build on the resonance pathway and improve the rubric approach.

The GPSA, other funders, and organizations working on social accountability should make intentional and long-term investments in robust research and evaluation initiatives using the relational rubric to continue to assess and test the resonance pathway to scale, based on the conceptions and preliminary evidence presented in this paper.²⁴

For example, to improve inter-rater reliability and replicability of the rubric method, multiple independent reviewers could assess the same projects to ensure that the core concepts and rubric level criteria are clear and well defined enough. Eventually the rubric could also provide an intermediary variable in measuring systemic-level social accountability and impact on public sector and civil society capacity, independent of the limited scope of a project. To test this, research or evaluation exercises designed to assess mixed results in the scale-up of social accountability processes and related impacts should be conducted. For example, looking into clusters (geographic, sector- and/or relational based), connected to the original project, where better public service delivery and development outcomes have materialized from new projects and policies, programs and policy reforms.

Understanding the complex dynamics at play and setting expectations accordingly.

Ensuring realistic expectations for the success of collaborative social accountability processes, recognizing the many legitimate forms of sustainability, the incremental steps involved, the inherent mixed results associated with complex processes, and long-term time frames required for scale-up.

²⁴ The authors credit and thank peer reviewer Mathieu Cloutier for these important recommendations on strengthening the replicability of the rubric method and enhancing its potential value and uses, through the investment in more targeted research and evaluation by social accountability funders and organizations.

Careful evaluator selection.

Independent evaluators need to bring the right technical skillset (methods) and a firm understanding of social accountability. External evaluations of social accountability interventions require a deep understanding and experience with social accountability work, as well as recognition of reasonable limits and expectations for sustainability. It is valuable if evaluators understand the nature of complex processes, such as those in a resonance pathway to scale (with social learning at its core), where scale-up happens through ongoing deliberation and negotiation between a wide set of actors, dynamics, and contextual shifts within a system. This will often result in non-linear processes, producing mixed results that do not look exactly like the original design, progressing in fits and starts over time.

Strong organizational commitment and investment.

Project and portfolio-level monitoring and evaluation using the relational rubric method requires competence and a good understanding of functional equivalents of core concepts. This requires a sound grounding of projects in a portfolio-level theory of change that can be applied across diverse individual project contexts. This requires intentional portfolio design; qualified and long-term MEL staff embedded within a program; a supportive leadership environment; and sufficient investment for systematic assessments repeated consistently over time.

In closing, the authors offer words of caution and advice for MEL and social accountability practitioners, and their respective organizations and funders. Applying these recommendations and the many enabling conditions can be challenging for MEL and social accountability practitioners, organizations, and funders. It requires the engagement of all stakeholders - from senior management to implementing staff in civil society organizations - to understand and commit to common concepts within a theory of change, and to monitor and evaluate a set of functional equivalent indicators from the project onset. This is necessary for reliable data and comparison over the long-term, aggregable at the portfolio level, and so that MEL systems can track cumulative results and impact. The nature of organizational restrictions, limited resources, technical criteria and continuously shifting political dynamics within organizations and the systems in which they work can make it difficult to embed these essential features. However, the findings in this paper are clear that the evidence and learning pay-off is worth the effort.

Annex A: Learning about Collaborative Social Accountability Sustainability and Scale-up through Quick Feedback Cycles

The conceptual framework for scale up with an eye towards sustainability and the relational rubric approach can apply across many projects and portfolios in the social accountability field. As explained in the main paper, sample of 15 projects funded by the World Bank's GPSA was used to test and illustrate the main arguments and findings.

The World Bank's decision to establish and fund a portfolio of locally tailored and led social accountability projects that collectively produce value for the field (see Box 7 below), is directly linked to the challenges and gaps in evidencing how social accountability can be sustained and scaled.

Box 7: The Creation of the GPSA and the World Bank's Mandate

In June 2012, the World Bank's Board established the GPSA to provide grants for CSO-led social accountability initiatives in partnership with governments, and to foster knowledge and learning about social accountability in different contexts. The imperative for scale, including as a mechanism to support sustainability, and the field's limited knowledge about it were both important considerations. The Board paper states:

“There is a need for more robust evidence on whether and how social accountability approaches can be sustained, scaled up, and replicated in different sociopolitical settings, and how international partnerships can leverage beneficial change. Addressing these knowledge gaps requires learning by doing, rather than passive research.”

The challenge set out by the board was about learning across the GPSA's portfolio, rather than from individual grants, and potentially connecting this learning with lessons from how scale happens across the broader social accountability field.

Source: World Bank data

According to the GPSA's Theory of Action:²⁵ the GPSA expects to demonstrate success when elements and lessons of small and experimental **“collaborative social accountability processes inform public sector decisions and actions beyond individual GPSA projects”** (GPSA, 2020). The term collaborative social accountability was used by the GPSA to distinguish different assumptions and forms of social accountability programming which exist in the field, acknowledging the diversity and its own comparative advantages. The specification of how this process might unfold put emphasis on the relational aspects of social accountability for short-, medium-, and long-term results, rather than on the tools and capacities that had been central to monitoring, research, and evaluation, but are only small parts of what social practitioners deploying collaborative approaches do or contribute to the deliberation, compromise and collective action that may contribute to scale with an eye towards sustainability.

By 2014, the GPSA had awarded two rounds of grants. It convened 165 global partners – a diverse group of development agencies, international and national CSOs, private sector groups and government representatives from around the world alongside World Bank staff. The GPSA Secretariat invited them to reflect on the scale of their interventions, capturing the discussion in a [GPSA learning note](#), which concluded:

25 For distinction between theory of action and theory of change, see Tyrrel, 2019.

“We discovered a range of on-the ground meanings for scale. We also uncovered that there may be some common challenges and similar pathways that show promise. A big ‘Aha!’ moment: in practice, our colleagues’ experience with scale is rarely directly associated with replication” (Guerzovich and Poli, 2014).

This insight was partly puzzling because, at the time, the replication of best practices for particular institutional forms across contexts²⁶ was widely assumed to be desirable and widely promoted in donor documents and governance approaches, as was noted in the World Development Report of 2017. Stakeholders’ fundraising strategies were telling donors what they seemed to want to hear. At the same time, the widespread assumption co-existed awkwardly with the mantra that ‘context-matters’. It also affected how similar interventions functioned in practice (Grandvoinet, Aslam and Raha, 2015), especially as the field had not identified the circumstances under which the replication of specific interventions might be more favorable.

For example, by 2014, the GPSA had received over 600 proposals for funding collaborative social accountability interventions. A systematic analysis of a sample revealed that only a few projects had a clear approach to scale. Guerzovich and Poli (2014) illustrate how a typical CSO applying for GPSA funding articulated its assumptions in this regard:

“[The CSO applying for GPSA funds proposed to] implement a pilot project in a range of local settings [it had] identified carefully and where [it would] work with local stakeholders to ensure adoption and implementation. Work in these areas of primary focus [would] help [the CSO] identify best practices that could be replicated elsewhere in the country. However, [the CSO] realize(d) that many of the key decisions about the process [it] care(d) about are made at the national level – i.e., not where [the CSO would be] working most of the time in [the proposed] project. Hence, [the CSO] would employ advocacy and awareness raising activities for national decision-makers taking advantage of the national networks the [CSO who applied] already belongs to. These networks [would] facilitate sharing of best practices and lessons learned to the wider national level audience and through the media for making a strong case for wider adoption of the model. The final phase of the knowledge and learning component of the project [would] focus on advocacy at the national level for country wide adoption of the model developed by the project. This process of wider dissemination and advocacy [would] contribute significantly to enhancing the knowledge base on local government dynamics, practices and intervention needs.”

Yet, early on, as the GPSA Secretariat’s capacity building team²⁷ began engaging and connecting civil society partners, World Bank teams and public officials, other approaches and entry points to growing impact began to look more promising and plausible. For example, in 2015, the formative evaluation of the Good Governance Practices for Dominican Republic project noted that public officials were advancing actions relevant to the project’s work (e.g., a transparency portal, the ‘Salir del Escondite’ campaign, among others) and this synergy could pay off in terms of project scale (Guerzovich, 2015). Similarly, a 2017 unpublished mid-term review of the Transparency and Accountability in Mongolia (TAME) project identified that investing in building synergies between a component of TAME and the World Bank-financed Education Quality Reform Project’s school grants component could offer a prospective pathway to scaling and sustaining insights from the TAME project. But to do so, CSO partners and World Bank teams had to listen to other perspectives, reflect on alternative scenarios available to them, and make a choice in terms of whether they would compromise their original vision towards scale up or not.

The GPSA team took note of these insights and began connecting the dots, using available resources to continue experimenting, gathering evidence and reflecting on additional knowledge from practice, and adaptive course-corrections. Moments to pause and reflect with grant project teams and annual grant partners’ meetings, among others, helped sharpen the focus of GPSA’s approach to scale. The parameters of this emergent approach to scale were integrated into specific projects, subsequent calls for proposals and reporting templates (Poli and Guerzovich, 2020a).

26 This pathway to scale “assumes that technical experts who produce knowledge can determine the unique form of social accountability mechanisms that work and then use their authority and knowledge to promote cross-context convergence (i.e., scale up) towards those arrangements” (Guerzovich et al 2022).

27 On the GPSA’s capacity building approach during this period see Poli and Guerzovich, 2020.

These emergent insights about how project teams build a road towards scale-up prospectively and emergently in practice, focused on function. This means *“identifying the systems and functions which need to be in place in order to support an ongoing process of state-citizen interaction around a particular problem or problem area.”* These diverged from some grant partners’ and evaluators’ expectations that the GPSA or a government would fund the continuity or expansion of a project’s exact form i.e., a technology, a tool, a standard.²⁸ Scale-up with an eye towards sustainability can be wholesale or partial happening gradually or immediately. Sometimes our expectations of transformative, wholesale change, mean that evaluators and practitioners fail to identify and document how more incremental change happens in practice over time (Guerzovich, 2022a). Scale-up can also be done effectively by others outside of government (e.g., CSOs/INGOs, donors, World Bank teams). In fact, grant partners later reflected that holding on to assumptions that CSOs would be funded on an ongoing basis to implement the same approaches in more sites or that their advocacy would enable wholesale uptake by national authorities could divert attention from more promising pathways to scale, and result in missed opportunities for scale-up. Furthermore, such misplaced, or unrealistic expectations often led to undue disappointment or a failure to celebrate successes that were happening in practice. This also perpetuated the misinformed narrative about social accountability failure regarding scale.

When the team reflected on the evaluation of Wahana Visi’s Citizen Voice and Action for Government Accountability and Improved Services: Maternal, Newborn, Infant and Child Health Services project in Indonesia, they confirmed that quick feedback loops are taking place at project and portfolio levels. Although this was insufficient to account for how scale happens in practice, the evaluation provided important insights to steer future GPSA evaluations towards an adaptive learning approach, as well as more generally for the GPSA’s monitoring and evaluation system for the whole program (see Box 8).

At the same time, the Indonesian evaluation focused on the role of civil society and, thus, did not go far enough. While it found that advocacy was not charting the pathway scale, it did not theorize nor look into how processes that were unfolding beyond civil society were triggering actions that could have done so. In Box 8 below, there is no straight line between civil society advocates and national authorities; rather multidirectional flows connect each actor’s work. The flows connecting actors entail give and take, social learning and collective action within and across sites, rather than replication of best practices or the exercise of civil society’s countervailing power. For instance, World Bank teams and documentation had complementary information about their work with government counterparts and other development partners which helped to spread lessons even wider. These lessons seem to have found their way into more policy dialogues and programming decisions than the evaluation suggested, including inspiring national officials to incentivize other public officials, funders and civil society groups to consider and adapt insights from the Wahana Visi project into their own work (Poli and Guerzovich, 2019; GPSA, 2016; authors interviews with stakeholders and GPSA documentation).

28 On the distinction between form and function, see Integrity Action (2020)

Box 8: Using Evaluation to Advance Knowledge About the Effects of Social Accountability

The evaluation of the GPSA-sponsored project in Indonesia delivered by grant partner Wahana Visi was the first in the social accountability field to systematically identify how collaborative social accountability can shift power asymmetries and strengthen health systems. Figure 7 below synthesizes the findings (World Bank Group, 2007). It was a breakthrough from traditional approaches to evaluating social accountability for at least three reasons:

1. It provided concrete evidence about the ways in which social accountability projects, which seemingly focused on producing more responsive service delivery, were producing systemic effects that are critical to support local systems strengthening and transformation after the end of a particular project. This viewpoint stands in contrast to widespread positions in the field that projects are limited to short-term effects, while long-term organic processes deliver long-term systemic transformations.²⁹
2. It helped GPSA partners and other organizations to highlight the work social accountability interventions are already doing to strengthen systems and produce more concrete knowledge about these systemic, but often implicit effects, on state-society relationships. When research and evaluations sought wholesale normative transformations of power relationships as well as short term production of transparency or other results, they omitted this important aspect of the work. It was unclear whether this was the result of an absence of evidence or evidence of absence. Conversely, when research and evaluations began explicitly asking about systemic effects, they found this effect which is valued by practitioners and communities alike. For example, J.B. Falisse and colleagues conducted independent evaluations of GPSA and non-GPSA projects in DRC. In the latter they reflected that:

“The real thread running through achievements ... is that the (collaborative social accountability) approach seems to allow the construction of a dialogue between parties that used to speak little (or not at all) to each other and an improvement in the relationship between the population, providers and the governmental side ... There are two ways (not necessarily opposed) to consider this renewed dialogue: either as a means to achieve the achievements described below or as an end in itself ... but let us emphasize here that dialogue is something that communities, providers, and authorities celebrate as an achievement in itself” (Falisse et al, nd; see also Falisse et al, 2019).

As stakeholders in the field spotted the blind spot, prioritized learning about these effects and began to ask questions about them, the absence of evidence and the possibility to address it became clearer.³⁰

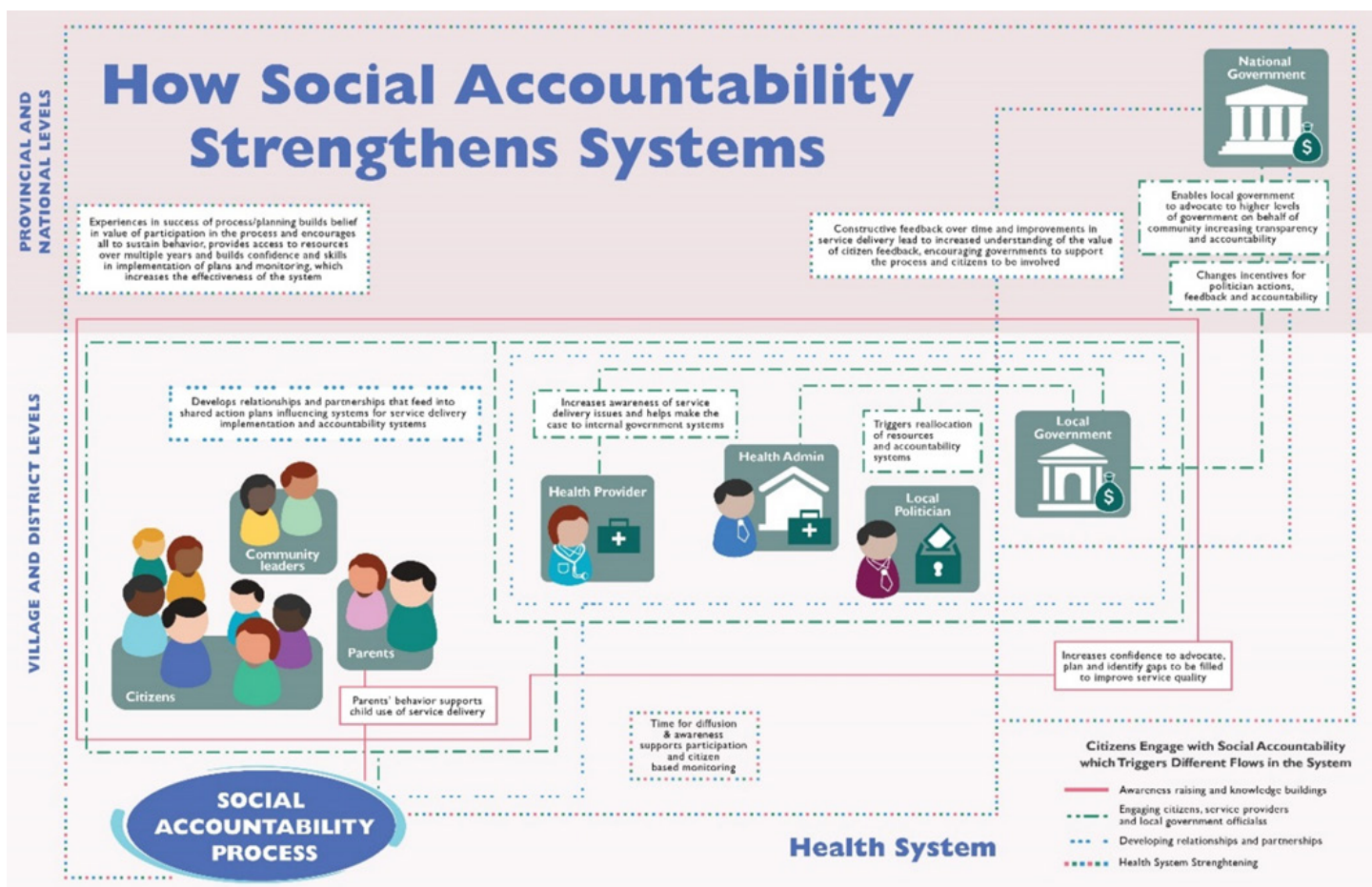
3. The Wahana Visi evaluation developed a methodology to trace and make causal claims about the concrete mechanisms which connect the facilities on the frontline to other sites of decision-making, informing collective action and social learning that stakeholders use beyond the frontline. The initial insights about these connections had more in common with the mid-term reviews of the GPSA Dominican Republic and Mongolia projects, than with the assumptions explored by the evaluation or those included in many of the funding proposals submitted to the GPSA.

Source: World Bank data

29 For a synthesis of these two positions, see Nelson et al, 2022.

30 For other evaluations that add to the body of evidence see Guertzovich, 2022.

Figure 7: The Creation of the GPSA and the World Bank’s Mandate



Source: Guertzovich 2022, adapted from Ball and Westthorp (2018)

Therefore, a theoretical gap still needed to be addressed by systematically exploring what seemed to be happening in Indonesia, Mongolia, and the Dominican Republic. Following the learning gained from the Wahana Visi evaluation, the GPSA team agreed with a series of grant partners to ask more direct questions about pathways to scale, encouraging evaluators to explicitly think prospectively about scale, looking beyond civil society’s direct action. Increasing investments in opening the black box could potentially help surface plausible connections between the GPSA’s project efforts and government actors and actions at different levels, as well as with the World Bank and other donors. The investment in using case-based causal analysis to conduct evaluations and produced inferences that could travel (under certain conditions) to other cases paid off.³¹ Evaluation findings further validated the assumption that something important was missing in the mainstream alternative theories of change about social accountability scale-up – whether it be those grounded in replication of best practice, those that betted on CSO’s adversarial countervailing power, or the hybrid reflected in applications for funding to the GPSA. The first evaluation of this set for the My School project in Moldova (*‘Scoala Mea’*) explicitly asked whether the project had influenced policy through the World Bank Group’s Country Management Office’s dialogue with the government. The evaluation found that *“the project provided information on World Bank operations in Moldova and the dialogue and strategies in the education sector to a relatively large extent”* (Costachi et al, 2018). Without asking and assessing this

31 On the rigor and potential of this use of cases for causal analysis see World Bank Group, 2023

directly, this vital finding and evidence for contribution to the project's scale (and therefore sustainability) might have gone unnoticed.

The GPSA team then shared the Moldovan findings with independent evaluators of other projects. Some accepted the challenge to use this new lens to inform their own assessments. Theoretical and methodological cross-fertilization across the portfolio helped uncover how change was happening across projects as well as the localization of the pathway in specific contexts. For example, the GPSA-funded public sector finance and budgeting project led by CSO SEND-Ghana was assessed as unlikely to be replicated in full in more districts due to financial constraints and competing prioritizations. However, the final evaluation identified three processes that were the *mostly likely* for the government to use (i.e., prospective sustainability): a multi-actor steering committee format to oversee citizen-public sector engagement on budgeting; an adapted spinoff of the project's dashboard for citizen engagement on public sector budgeting;³² and the project's learning and insights to help monitor other World Bank projects. It also identified the government and its parliament's new expectation and demand for SEND-Ghana's future technical inputs into these processes – all of which require give and take from those involved (Mills, 2019). Furthermore, in the DRC, the GPSA-funded health sector project led by Cordaid informed other development partners' programming. This was facilitated by local citizens who are redeploying collective action learned from the GPSA project towards other efforts to improve local health services delivery by working with local government and other relevant stakeholders, listening to multiple perspectives, and reaching adaptive compromises to move forward (World Bank Group, 2020).

The evaluations of the Moldova, Ghana and DRC projects were uncovering significant but previously ignored results. Collectively, these evaluations provided new, valuable data to specify how collaborative social accountability operated and which results should be prioritized to better understand and inform approaches for evidencing sustainability and scale-up. Consequently, emerging insights informed the next iteration of the GPSA's Theory of Action and its monitoring, evaluation, and learning processes. At the same time, the GPSA also revamped its Results Framework outcomes and indicators and began developing fit-for-purpose MEL approaches that could help it to better understand how the results of individual projects 'added up' and contributed to wider impact and learning for the social accountability field. This was also necessary to test the assumptions in its Theory of Action and validate the logic of the expected short to long-term results, at the program level. This challenge is discussed in Box 9 below.

Encouragingly, these initial findings were further validated by subsequent GPSA project evaluations. For example, the final evaluation of the Improved Social Accountability for Bettering Preschool Quality in Georgia project made a positive assessment of its interventions' sustainability. Tangible results were a combination of strong ownership (including by stakeholders in municipalities); direct references in the Government's draft Education Strategy; the transfer of learning to a new World Bank operation; and that civil society partners now had a seat at the table of the country's education governance.³³

32 The Dashboard was an ICT interactive social accountability platform for citizens to report or raise concerns with state actors at the district and national levels. The intention was to enable citizen to give the government feedback in real time.

33 "From the perspective of social accountability, the intangible results of strengthening relationships, experiences gained in collaborative action, and improved agency on the part of project beneficiaries and local stakeholders, are likely to continue after the project completion." (Ecorys, 2020)

Box 9: Greater Than the Sum of Its Parts? The Challenge of Evidencing Impact at the Portfolio Level

The GPSA's grantmaking instrument funds individual projects for civil society organizations to lead responses to local public service delivery and policy problems that may benefit from collaborative problem-solving amongst citizens, civil society, and the public sector. The GPSA invested US\$50 million (and over \$6 million in parallel funding) in a portfolio of 51 grants for CSO-led collaborative social accountability initiatives in 34 countries and a range of public service sectors such as health, education, governance, social protection, water, agriculture, and public finance, between 2012 and 2022. As per its Theory of Action, *"The nature of the GPSA's grant-making is to make small experimental investments with the potential for scale-up and sustainability. When elements and lessons of collaborative social accountability processes inform public sector decisions and actions beyond individual GPSA projects, the GPSA demonstrates success."*

A portfolio-based approach like this one also acts as a ripe platform for strategic learning for action across these complementary experimental interventions. The GPSA was set up with portfolio-level ambitions namely contributing to field learning about collaborative social accountability. However, as all projects are also localized to fit their unique context and relatively small scale in nature, it can be challenging to compare and aggregate evidence at this higher level, to understand whether the combined efforts of all the projects contribute to something that is greater than the sum of their parts. Comparing diverse results and aggregating them in a fit-for-purpose and meaningful way is difficult, especially considering all the diverse forms that scale and sustainability can take, as well as mixed results.

This challenge is not unique to the GPSA - many funders and organizations in the transparency, participation and accountability sector (and other development fields) have long been grappling with this challenge. As more organizations are focused on complex, system-level transformations, more [civil society organizations](#) and [funders](#) are seeking ways to solve this evidence and evidencing gap. The GPSA has been tackling this over the past few years with several connected efforts including a revised Theory of Action and Results Framework of 2020 and moving towards a harmonized Monitoring, Evaluation, Reporting and Learning (MERL) System with specific indicators and methods for projects to use and feed into their results and learning to enable portfolio level analysis, including additive effects in terms of the development of the system. This includes the approach discussed in this note about evidencing scale, sustainability, and uptake of collaborative social accountability processes.

Source: Own elaboration

Despite these advances in learning and evidence about scale-up, a few following key related questions remained. Ways to address these are discussed in the main paper.

- i) Could the insights from these and other GPSA projects produce lessons about a whole that was more than the sum of the parts (see Box 9 above)?
- ii) How did this Theory of Action fit into a broader understanding of how change happens within collaborative social accountability processes?
- iii) Is this emerging understanding also applicable to explain the results of the swath of other social accountability programming?³⁴
- iv) If so, could the GPSA be on track to produce and consolidate learning by doing that delivered on the World Bank's Board mandate to contribute towards *“more robust evidence on whether and how social accountability approaches can be sustained, scaled up, and replicated in different sociopolitical settings, and how international partnerships can leverage beneficial change”*?



34 This paper explores how new data informed the GPSA's work. Insights from GPSA team members and partners are cross-fertilized through working with other organizations and partners. Arguably then the GPSA's modest investments in thought leadership, through writing, as well as convening events spill overed beyond its own programing. However, tracing that contribution is beyond the scope of this paper. For examples of cross-fertilization beyond the GPSA portfolio, see (Jacobstein, 2022; Guertzovich and Gondo, 2022; Guertzovich et al, 2017; Guertzovich, 2022c)

08

—

References

- Andrews, Matt; Pritchett, Lant; Woolcock, Michael. 2017. *Looking like a state: The seduction of isomorphic mimicry*. <https://academic.oup.com/book/26994/chapter/196206819>
- Aston, Thomas. 2022. "Introducing a Resonance Pathway to Scale" published 4 January, 2022, accessed at <https://thomasmtaston.medium.com/introducing-a-resonance-pathway-to-scale-6cacad5163cd8>
- Aston, Thomas. 2020. *Rubrics as a harness for complexity*. <https://thomasmtaston.medium.com/rubrics-as-a-harness-for-complexity-6507b36f312e>
- Aston, Thomas; Guerzovich, Florencia; and Wadeson, Alix. 2021. "Tales of triumph and disaster in the transparency, participation, and accountability sector." published 26 August 2021 accessed at: <https://thomasmtaston.medium.com/tales-of-triumph-and-disaster-in-the-transparency-participation-and-accountability-sector-5f638261983c>
- Aston, T.; & Zimmer Santos, G. (2022). *Social Accountability and Service Delivery Effectiveness: What is the Evidence for the Role of Sanctions*. GPSA
- Bovens, M.; Goodin, R. E., Schillemans, T., Bovens, M; Schillemans, T; and Goodin, R. E. 2014. Public Accountability. In *The Oxford Handbook of Public Accountability*. <https://doi.org/10.1093/oxfordhb/9780199641253.013.0012>
- Chingaipe, Henry; Thombozi, Joseph; Katundu, Enea and Bongololo, Grace. 2022. *Evaluation of the GPSA Program Based on Projects in the Primary Education Sector in Malawi*.
- Cloutier, Mathieu. 2021. *Social Contracts in Sub-Saharan Africa: Concepts and Measurements*. World Bank Group. Governance Global Practice. Policy Research Working Paper 9788
- Collier, D. and Mahon, JE. 1993. "Conceptual stretching revisited: adapting categories in comparative analysis." *Am Polit Sci Rev* 87(4):845–855.
- Costachi, Ionela; Criste, Aliona; Terzi Barbarosie, Daniela. 2018. *My School - Empowered Citizens Enhancing Accountability of the Education Reform and Quality of Education in Moldova (English)*. Washington, D.C.: World Bank Group, p.44. <http://documents.worldbank.org/curated/en/744681621832061746/My-School-Empowered-Citizens-Enhancing-Accountability-of-the-Education-Reform-and-Quality-of-Education-in-Moldova>
- Doin, Guilherme Augusto; Dahmer, Jeferson; Schommer, Paula Chies; Spaniol, Enio Luiz. 2012. *Mobilização Social E Coprodução Do Controle: O Que Sinalizam Os Processos De Construção Da Lei Da Ficha Limpa E Da Rede Observatório Social Do Brasil De Controle Social*
- Ecorys, 2020. *Reinforcing Social Accountability in Health Services in Sud Kivu and Kongo Central Provinces: Final Evaluation of the GPSA-CODESA Project (English)*. Washington, D.C.: World Bank Group. <http://documents.worldbank.org/curated/en/883991607354429777/Final-Evaluation-of-the-GPSA-CODESA-Project>
- E-Pact Consortium. 2016. *Empowerment and Accountability Annual Technical Report 2016: What Works for Social Accountability, Macro Evaluation of DFID's Policy Frame for Empowerment and Accountability*. Oxford: e-PACT. Fox, J. 2014.
- Falleti, T. G., & Mahoney, J. 2015. "The comparative sequential method." In J. Mahoney, & K. Thelen (Eds.), *Advances in Comparative-Historical Analysis (Strategies for Social Inquiry)* (1st edition ed., pp. 211-239). Cambridge University Press.
- Falisse, Jean-Benoît; Mulongo, Philémon; and Koko Kirusha, Janvier. n.d. "The Voice and Citizen Action (CVA) Approach of World Vision DRC Meta-Evaluation (2013-2020)." World Vision International.
- Falisse, J., Mafuta, E., & Mulongo, P. 2019. *Reinforcing Social Accountability in Health Services in Sud Kivu and Kongo Central Provinces. Final Evaluation of the GPSA/CODESA Project*.
- Fox, Jonathan. 2014. *Social Accountability: What does the evidence really say?* GPSA Working Paper No. 1, Washington, DC: International Bank for Reconstruction and Development.
- Global Partnership for Social Accountability. 2022. *Scaling Social Accountability to Create Lasting Change* <https://thegpsa.org/sessions/scaling-social-accountability-to-create-lasting-change/>

Scaling Up Collaborative Social Accountability in Complex Governance Systems:
A Relational Approach for Evidencing Sustainability

- Global Partnership for Social Accountability. 2021a. *How to make mid-level theory more useful for social accountability that contributes to building back better?* <https://vimeo.com/558157200>
- Global Partnership for Social Accountability. 2021. *Global Partners Forum. Social Accountability for a Strong COVID-19 Recovery.* <https://vimeo.com/564191691>
- Global Partnership for Social Accountability. 2020. *Theory of Action.* <https://documents1.worldbank.org/curated/en/425301607358292998/pdf/The-Global-Partnership-for-Social-Accountability-Theory-of-Action.pdf>
- Global Partnerships for Social Accountability. 2019a. *Global Partners Forum. Social Accountability and the Challenge of Inclusion.* <https://vimeo.com/451986935?signup=true>
- Global Partnership for Social Accountability. 2019. *Improving Transparency and Performance of the Conditional Cash Transfer Program.* GPSA project. <https://thegpsa.org/projects/improving-transparency-and-performance-of-the-conditional-cash-transfer-program/>
- Global Partnership for Social Accountability. 2018. *Improving Social Accountability in the Water Sector Through the Development of Quality Standards and Citizen Participation in Monitoring in Tajikistan.* GPSA Project. <https://thegpsa.org/projects/improving-social-accountability-in-the-water-sector-through-the-development-of-quality-standards-and-citizen-participation-in-monitoring-in-tajikistan/>
- Grandvoinet, Helene; Aslam, Ghazia; Raha, Shomikho. 2015. *Opening the Black Box: The Contextual Drives of Social Accountability.* <https://elibrary.worldbank.org/doi/abs/10.1596/978-1-4648-0481-6>
- Guerzovich, Florencia. 2023a. "Does the whole add more than the SUM of its parts?" <https://medium.com/@florcig/does-the-whole-add-more-than-the-sum-of-its-parts-8b9eb352bb67>
- Guerzovich, Florencia. 2023. "Adaptive Management Across Project Cycles: Look into Coherence in Time" <https://medium.com/@florcig/adaptive-management-across-project-cycles-look-into-coherence-in-time-ab99caa3a9e5>
- Guerzovich, Florencia. 2022. *Systems-Aware Social Accountability (SASA): Supporting the Whole to Be Greater than the Sum of Its Parts.* Pact. Washington, DC.
- Guerzovich, Florencia. 2022b. "How Context Shapes Pathways to Scale in Social Accountability". <https://medium.com/@florcig/how-context-shapes-pathways-to-scale-in-social-accountability-post-4-of-5-d417cfe2b4f5>
- Guerzovich, Florencia. 2022a. "Scale up In Time: Revisiting How we Evidence Process and Context." <https://medium.com/@florcig/scale-up-in-time-revisiting-how-we-evidence-process-context-6c53f82a1817>
- Guerzovich, Florencia. 2022c. *Literature Review towards a WV theory on Social Accountability as a Driver of Sustainable Child Well-Being.* World Vision International.
- Guerzovich, Florencia. 2021c. "Pathways to scale in social accountability." Published December 21, 2023. <https://medium.com/@florcig/pathways-to-scale-in-social-accountability-post-1-of-5-40e5ff51a053>
- Guerzovich, Florencia. 2021b. "Learning from consortia and portfolios: From cacophony to symphony" published August 31, 2021, accessed at: <https://medium.com/@florcig/learning-from-consortia-and-portfolios-from-cacophony-to-symphony-ab0c8ddedaff>
- Guerzovich, Florencia. 2015. *Evaluación Formativa Proyecto Prácticas de Buen Gobierno en República Dominicana (Proyecto Vigilantes)* <https://documents1.worldbank.org/curated/en/900101607357369855/pdf/Evaluacion-Formativa-Proyecto-Practicas-de-Buen-Gobierno-en-Republica-Dominicana-Proyecto-Vigilantes.pdf>
- Guerzovich, Maria F. and Aston, Tom. 2023, "Social Accountability 3.0: Engaging Citizens to Increase Systemic Responsiveness" (July 19, 2023). Available at SSRN: <https://ssrn.com/abstract=4606929>
- Guerzovich, Maria F; Aston, Tom; Levy, Brian; Chies Schommer, Paula; Haines, Rebecca; Cant, Sue; Faria Zimmer Santos, Grazielli. 2022. *How do we shape and navigate pathways to social accountability scale? Introducing a middle-level Theory of Change,* CEDIL Research Project Paper 1. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford. <https://policycommons.net/artifacts/3533985/cedil-research-project-paper-1/4335198/>

- Guerzovich, Florencia and Gondo, Rachel. 2022. "Social Accountability Practitioners as System Conveners." <https://Medium.Com/@florcig/Social-Accountability-Practitioners-as-System-Conveners-33b77c8a4778>
- Guerzovich, Florencia; Yeukai Mukorombindo; and Elsie Eyakuze. 2017. "Beyond Fundamentals: Learning About Social Accountability Monitoring Capacities and Action in Southern Africa." PSAM. Grahamstown.
- Guerzoich, Florencia; Poli, Maria. 2020. *How Social Accountability Strengthens Cross-sector Initiatives to Deliver Quality Health Service?* GPSA Note 17 Washington, D.C. World Bank Group. <https://documents1.worldbank.org/curated/en/600891606911830725/pdf/How-Social-Accountability-Strengthens-Cross-Sector-Initiatives-to-Deliver-Quality-Health-Services.pdf>
- Green, Duncan. 2017. *Theories of Change for Promoting Empowerment and Accountability in Fragile and Conflict-Affected Settings*. IDS Working Paper 499. <https://www.ids.ac.uk/publications/theories-of-change-for-promoting-empowerment-and-accountability-in-fragile-and-conflict-affected-settings/>
- Guerzovich, Maria Florencia; Poli, Maria. 2014. *How are GPSA's Partners Thinking About Scale and Trying to Achieve It (English)*. Global Partnership for Social Accountability (GPSA), Note No. 8 Washington, D.C. World Bank Group. <http://documents.worldbank.org/curated/en/654161606890404650/How-are-GPSA-s-Partners-Thinking-About-Scale-and-Trying-to-Achieve-It>
- GPSA. 2016. *OGP Summit Workshop, Coproducing Open Government Results: Insights from the Global Partnership for Social Accountability*. <https://thegpsa.org/event/ogp-summit-workshop-co-producing-open-government-results-insights-from-the-global-partnership-for-social-accountability/>
- Haldrup, Soren Vester. 2020. "Measuring Systems Transformation: Towards a Preliminary Framework". UNDP Strategic Innovation. Accessed at <https://medium.com/@undp.innovation/measuring-systems-transformation-towards-a-preliminary-framework-958ad3444949>
- IEG. 2017. *Rethinking Evaluation. Reflections from Caroline Heider*. <https://ieg.worldbankgroup.org/sites/default/files/Data/RethinkingEvaluation.pdf>
- Integrity Action. 2020. "Citizen-Centred Accountability: How Can We Make it Last?" Briefing Note. October 2020. Accessed at: https://integrityaction.org/media/16127/integrity-action-sustainability-research-briefing-note_.pdf
- Jacobstein, 2020. *What is the Work?* <https://usaidlearninglab.org/community/blog/what-work>
- Jacobstein, David. 2019. *Market Systems Insights for DRG - Success as a Dynamic System*. <https://usaidlearninglab.org/community/blog/market-systems-insights-drq-success-dynamic-system>
- Jespersen, Ann-Sofie. 2022. *Civil Society Actions Push Reforms in Mauritanian Schools*. <https://www.globalpartnership.org/blog/civil-society-actions-push-reforms-mauritanian-schools>
- Kania, John; Kramer, Mark; and Senge, Peter. 2018. *The Water of Systems Change* FSG
- KOMPAK (Kolaborasi Masyarakat dan Pelayanan untuk Kesejahteraan). 2018. "KOMPAK Program Logic and Ways of Working 2018–2022." Jakarta, Indonesia: KOMPAK. <https://www.dfat.gov.au/sites/default/files/indonesia-kompak-program-logic-and-ways-of-working-2018-2022.pdf>
- Lekweiry, Mohamedou Ould and Falisse, Jean-Benoit. 2022. *Final Evaluation of the Transparency of the Mauritanian Education Budget (TOME) Project (English)*. Washington, D.C.: World Bank Group. <http://documents.worldbank.org/curated/en/352401647610255802/Final-Evaluation-of-the-Transparency-of-the-Mauritanian-Education-Budget-TOME-Project>
- Meadwell, Hudson. 2022. *Endogeneity and qualitative political analysis: Debates about method or debates about ontology?* <https://journals.sagepub.com/doi/full/10.1177/05390184221138493>
- Meyanathan, Saha. 2021. *Catalysts for Change: Parent-Teacher Association in Mongolian Schools. Final Evaluation of the Transparency and Accountability in Mongolian Education Project*.
- Mills, Linnea Cecilia. 2019. *Making the Budget Work for Ghana: Final Evaluation (English)*. Washington, D.C.: World Bank Group. <http://documents.worldbank.org/curated/en/418131607356580690/Making-the-Budget-Work-for-Ghana-Final-Evaluation>

- Nelson, E., Waiswa, P., Coelho, V. S., & Sarriot, E. 2022. Social accountability and health systems' change, beyond the shock of Covid-19: drawing on histories of technical and activist approaches to rethink a shared code of practice. *International Journal for Equity in Health*, 21(S1), 41. <https://doi.org/10.1186/s12939-022-01645-0>
- OECD. 2021. "Using the evaluation criteria in practice", in *Applying Evaluation Criteria Thoughtfully*. OECD Publishing, Paris, https://www.oecd-ilibrary.org/development/applying-evaluation-criteria-thoughtfully_d1aca6d0-en
- OECD DAC. 2019. *Evaluation Criteria* <https://www.oecd.org/dac/evaluation/dacriteriaforevaluatingdevelopmentassistance.htm>
- Ostrom E. 1990. *Governing the commons: The evolution of institutions for collective action*. Cambridge, UK: Cambridge University Press.
- Patton, Michael Quinn. 2020. *Evaluation Criteria for Evaluating Transformation: Implications for the Coronavirus Pandemic and the Global Climate Emergency*. American Journal of Evaluation. Volume 42. Issue 1. <https://doi.org/10.1177/109821402093368>
- Pierson, P. 2004. *Politics in Time: History, Institutions, and Social Analysis*.
- Poli, Maria; Guerzovich, Maria Florencia. 2020a. *Integrating Adaptive Learning in Grant-Making: The Case of the GPSA (English)*. Global Partnership for Social Accountability (GPSA), Note No.16 Washington, D.C.: World Bank Group. <http://documents.worldbank.org/curated/en/116071606910702575/Integrating-Adaptive-Learning-in-Grant-Making-The-Case-of-the-GPSA>
- Poli, Maria; and Guerzovich, Maria Florencia. 2020. *Capacity and Implementation Support Area: Portfolio Performance Review (English)*. Global Partnership for Social Accountability (GPSA), Note No.15 Washington, D.C.: World Bank Group. <http://documents.worldbank.org/curated/en/893741606909911810/Capacity-and-Implementation-Support-Area-Portfolio-Performance-Review>.
- Poli, Maria; Guerzovich, Maria Florencia. 2019. *How Social Accountability Strengthens Cross-Sector Initiatives to Deliver Quality Health Services (English)*. Global Partnership for Social Accountability (GPSA), Note No.17 Washington, D.C.: World Bank Group.
- Tyrrel, Lavinia. 2019. *Theory of Change and Theory of Action: What's the difference and why does it matter?* <https://abtgovernance.com/2019/07/19/theory-of-change-and-theory-of-action-whats-the-difference-and-why-does-it-matter/>
- Wadeson, Alix; and Guerzovich, Florencia. 2023. *Monitoring, Evaluation, Reporting and Learning Guide for GPSA Grant Partners and Consultants*. World Bank, Washington, DC.
- Wadeson, Alix. 2022. Internal World Bank Document.
- Wadeson, Alix. 2021. "Orchestrating a MEL system for portfolios and programs: what we're testing now." <https://medium.com/@alixsara/orchestrating-a-mel-system-for-portfolios-and-programs-what-were-testing-now-4ca76210c2b7>
- Wadeson, Alix; Monzani, Bernardo; Aston, Tom. 2020. *Process Tracing as a Practical Evaluation Method: Comparative Learning from Six Evaluations*. https://mande.co.uk/wp-content/uploads/2020/03/Process-Tracing-as-a-Practical-Evaluation-Method_23March-Final.pdf
- Wenger-Trayner, Beverly. 2014. *What is Social Learning?* <https://www.wenger-trayner.com/what-is-social-learning/#:~:text=Social%20learning%20in%20the%20way,in%20something%20they%20care%20about>
- Wenger-Trayner, E; and Wenger-Trayner B. 2021. *Systems converging: a crucial form of leadership for the 21st century*. Social Learning Lab.
- World Bank Group. 2023. *The Rigor of Case-Based Causal Analysis Busting Myths through a Demonstration*. [The Rigor of Case-Based Causal Analysis \(worldbankgroup.org\)](https://www.worldbank.org/en/publication/the-rigor-of-case-based-causal-analysis)
- World Bank Group. 2020. *Reinforcing Social Accountability in Health Services in Sud Kivu and Kongo Central Provinces: Final Evaluation of the GPSA-CODESA Project (English)*. Washington, D.C.: World Bank Group. <http://documents.worldbank.org/curated/en/883991607354429777/Final-Evaluation-of-the-GPSA-CODESA-Project>

World Bank Group. 2017. *2017: Governance World Development Report and the Law*. Washington, DC: World Bank. © World Bank. <https://openknowledge.worldbank.org/handle/10986/25880> License: CC BY 3.0 IGO.

World Bank Group. 2007. *Citizen Voice and Action for Government Accountability and Improved Services: Maternal, Newborn, Infant and Child Health Services: Final Evaluation Report (English)*. Washington, D.C.: World Bank Group. <http://documents.worldbank.org/curated/en/331651607355519716/Final-Evaluation-Report>