

**Global Partnership for Social Accountability (GPSA)  
Results Framework Operational Indicators Retroactive Coding Exercise:  
Findings and Process Report  
25 February 2022**

---

## 1. Introduction

The GPSA has developed and continues to test and iterate a Results Framework for the program with a set of targeted outcomes and outputs. This includes a set of both grant partner and World Bank indicators linked to the most recent version of the GPSA Theory of Action. In late 2020, a consultant was hired to support the methodology and operationalization of Monitoring, Evaluation, Reporting and Learning (MERL). A comprehensive MERL guide has been developed (in process of finalization at the time of this report) with specific methodology and guidance for assessing each of the relevant indicators to be monitored and evaluated for all grant partner projects under the GPSA. These are referred to as 'operational indicators' and are the focus of this exercise.



*Figure 1: Timeline of the GPSA MERL process*

In 2021-2022, the GPSA also embarked on a Strategic Review Process (SRP) and Forward Look to better understand its progress and performance against the Theory of Action and the Results Framework, amongst other elements. As a part of this exercise, the GPSA decided to retroactively assess a sample of projects to date under the GPSA since the first call for proposals in 2013 using the updated Results Framework and MERL guide methodology.

This report provides an overview of the process and findings of this retroactive coding exercise. It is intended to inform part of the SRP/Forward Look and support the GPSA in its overall assessment and decision-making process. It also provides an important institutional memory for others who may need to reference and understand how the process for the retroactive coding took place as well as observations on the dataset and exercise overall.

It is important to note that there is an accompanying reference document (in excel) with the full set of sample coded projects as well as the aggregate values that contains more explanation, source evidence and important qualitative details about the findings and limitations for each project in the sample. It should also be understood the wider context of the GPSA Theory of Action and the MERL Guide document which is the key detailed reference document for how and why the Results Framework indicators should be understood, designed, assessment, monitored, reported and evaluated for all GPSA projects.

The exercise demonstrated how the Theory of Change and Results Framework can be evidenced in practice, indicating that a) GPSA projects can be assessed against the Theory of Action and Results Framework through the methods outline in the MERL Guide, using even secondary data (with limitations) and a) the results of the coding projects show the GPSA projects are realizing results in line with the trajectory of the GPSA Theory of Action and the intended outcomes of the Results Framework. More

consistent use of the methods and standardization of project and portfolio level assessment process over time will be needed to build on this preliminary evidence. However, as a first exercise, it reflects the promising potential for the GPSA in terms of evidencing the strategic and practical value of GPSA's investment in small-scale civil society-led collaborative social accountability grants to target proximate causes of service delivery failure and jointly problem solve with public sector and community counterparts in a diverse range of public sectors and countries.

## 2. Process Overview

This section provides an overview of the process of how the exercise was designed and executed including the objectives, scope, sample selection, methodology and limitations.

### 2.1 Objectives

- To support the SRP/Forward Look by coding and aggregating data from a sample of GPSA projects against the updated GPSA Results Framework's operational indicators.
- To test the MERL guide criteria for scoring and coding Results Framework operational indicators (noting the limitations of using a secondary limited dataset).
- To provide concrete examples and source materials on how to code the GPSA RF indicators for other stakeholders (MERL Consultants, GPSA/World Bank staff, grant partners)
- To provide a detailed reference document for GPSA institutional memory (i.e., the fully coded and aggregated sample set and the process and findings report)

### 2.2 Scope

This coding exercise focused on the operational indicators from the GPSA Results Framework which are intended to be included in all grant partner project Results Framework (or a fit-for-purpose functional equivalent). These are provided below for reference. The other indicators meant for World Bank/GPSA performance assessment are not included with the exception of output 3 indicator specifically. While this is not for grant partner purposes, it was deemed as possible to retroactively code for this exercise and was therefore included. Therefore, when reviewing the below Results Framework, it is important to remember that it does not include all the indicators. For reference to all indicators, please see the complete Results Framework on the GPSA website.

Table 1: Adapted version of the GPSA Results Framework for indicators included in the coding exercise

GPSA Results Framework outcomes/outputs	Operational Indicators
<p><i>Outcome 1: Civil society partners (lead grant partner and implementing CSO partners) have improved capacity to engage meaningfully and collaboratively in policy making and implementation and service delivery processes.</i></p>	<p>1a) Percentage of grant partners and their implementing CSO partners with improved capacity to engage meaningfully and collaboratively with one another and other stakeholders in the compact(s).</p> <p>1b) Percentage of grants in which civil society actors, citizens and public sector institutions demonstrate improved capacity to engage meaningfully and collaboratively.</p> <p>1c) Percentage of grants that contribute to corrective measures taken by public sector actors to address proximate causes of service</p>

	<p>delivery failure.</p> <p>The third indicator (1c) is a measure of the grant's contribution to improvement in service delivery in each project, based on the requirements of each grant. This indicator will be at the Project Development Objective level in grant partner projects.</p>
<p><i>Outcome 2: Civil society partnerships (lead grant partner and CSO implementing partners) and relevant public sector counterparts engage in collaborative social accountability processes that include citizens.</i></p>	<p>2) Percentage of GPSA grants in which civil society partnerships and relevant, targeted public sector counterparts engage in collaborative social accountability processes that include citizens.</p>
<p><i>Outcome 3: Collaborative social accountability processes are used to target proximate causes of service delivery failure to improve targeted service delivery.</i></p>	<p>3) Percentage of GPSA grants in which collaborative social accountability processes targeted the proximate cause of service delivery failure.</p>
<p><i>Outcome 4: Elements of collaborative social accountability processes are taken up by governments beyond individual GPSA projects.</i></p>	<p>4a) Percentage of GPSA grants in which governments seek to:</p> <ul style="list-style-type: none"> <li>(i) Use substantive lessons for improvements of targeted policies, processes, and mechanisms;</li> <li>(ii) Sustain elements of collaborative social accountability processes after the life of the project;</li> <li>(iii) Adapt insights from GPSA projects to scale them through programs or policies; or</li> <li>(iv) Apply elements of collaborative social accountability processes in additional localities or sectors.</li> </ul> <p>Note: this can be done through the government's own reform program, donor-funded programs, or Bank-financed programs.</p> <p>[Target: 25%]</p> <p>indicator is not for inclusion in project-level Results Frameworks.</p>
<p><i>Output 1: Civil society GPSA grant partners lead multi-stakeholder compacts.</i></p>	<p>1a) Percentage of GPSA grants with partner-led compacts with involvement from at least 3 fit-for-purpose stakeholder groups</p>
<p><i>Output 2: Lessons from experience inform GPSA engagement.</i></p>	<p>2a) Percentage of GPSA of grants in which lessons learned during implementation informed course corrections.</p>

<p><i>Output 3: World Bank sector teams support meaningful engagement between civil society and government.</i></p>	<p>3) Number of grants for which World Bank sector teams supported engagement between civil society and government.</p>
---	---

### 2.3 Sample and selection process

There has been a total of 35 grant partner projects funded by the GPSA across the four calls for proposals since 2013 at the time of designing this retroactive coding exercise. However not all the projects were included in the sample due to two key reasons:

- The available secondary documentary evidence from each project. To enable coding against the coding frame of the MERL guide for each indicator, especially relying on only secondary data, there needs to be a minimum amount of documents across each project life (grant partner reports, learning documents, analyses, independent reviews and evaluations and the World Bank Implementation Status and Results Reports (ISR) and Implementation Completion and Results Reports (ICR)). The consultant reviewed each the set of available project files to assess which ones had a minimum that would enable them to code the project against the GPSA Results Framework indicators that would yield reliable and useful information and to also warrant the time/budget resources involved in doing so. Several projects did not meet the criteria and were therefore excluded – see table below for more information.
- Several projects were too new into their start-up and implementation to have generated reporting or other documentation that would enable the coding to be done. In these cases, the projects were excluded – see table below for more information.

A total of **15 projects fulfilled the criteria for the coding exercise and were therefore included** in the sample. The table below provides the details on all the included and excluded projects in the sample frame and the reasons for exclusion of 20 projects. However, it is important to consider that because 9 projects fall into the ‘too early into implementation’ category, the total of projects that could have been possibly coded falls to 26. Therefore, a sample of 15 represents over half of these projects, or 62.5%.

*Table 2: Sample of GPSA projects for the retroactive e coding exercise*

### 2.4 Methodology

The method for coding each operational indicator used the criteria and instructions in the MERL Guide to the closest extent possible with the available secondary data. Due to the limitations (see below), there inherent gaps and trade-offs made to enable the coding to work. However, the 15 sample projects and process provided enough rigor and relevant proxy data to assign values to all the indicators for all the projects in the sample and to aggregate the data. The MERL Guide and methods for the Results Framework indicators is new and are being tested and therefore still a work in progress. This retroactive coding provided a good initial ‘test’, despite the limitations. For information on the fully designed method for each indicator (as intended with primary and secondary data), please refer to the MERL Guide.

### 2.5 Limitations

**Documentation and data quality:** The evidence available for the retroactive coding exercise was limited to the available secondary documents of each GPSA project in the sample. These varied by project in terms of number of documents, type of documents and quality of documents. There were some projects with robust reporting across the project life and higher quality ISRs/ICRs and independent evaluations while others had significant gaps (see section 4 for more details on data quality). In particular, the Results Frameworks reviewed in the ICRs often lacked the requisite qualitative narratives to substantiate and explain the scores. For the GPSA, the quantitative indicator values in Results Framework lack meaning and rigour without the accompanying causal analysis and qualitative narrative to substantiate them. Furthermore, many

evaluations reviewed did not specifically assess project Results Framework indicators, making it unfeasible to verify the indicator values provided in the ICRs.

**Indicators:** As many projects in the sample were not using the same indicators in the GPSA Results Framework, the consultant did a mapping exercise to match the most relevant indicators to the GPSA ones. These were used to inform the GPSA coded values along with the documentary evidence reviewed. However, as the indicators were not the same, units of analysis and values were not 'like for like' when compared to the GPSA indicators (they were not functional equivalents). Additionally, the coder could not verify how Results Framework values in the grant partner final Results Framework scores (as shown in the ICR) were calculated. In many cases, the qualitative sections of the Results Framework ICR were missing significant details to draw any substantial conclusions; therefore, the coder had to take them at face value.

While in some cases the GPSA Secretariat team could have provided additional information to revisit certain coding to triangulate and add tacit evidence, due to the targeted nature of GPSA Secretariat engagement and the team's availability, the same institutional memory/knowledge was not available across the sample of cases. In many cases this means that some projects do not benefit from additional information that could have supported higher or lower values. The decision to use only project documentation systematically across the portfolio was prioritized over maximizing the evidence base for individual projects because the goal of this exercise is to produce a broader picture of the portfolio as a whole. However, in moving forward, the documented regular exchanges between the GPSA, TTLs and grant partners will be a useful source of evidence to draw evidence for assessing Results Framework indicators.

**Sample size:** It is a relatively small sample set, however as mentioned above, in the universe of funded grant partner GPSA projects since 2013 (35) and those which were not too early in implementation (26), the sample still represents over half of the available projects for coding.

**Inter-rater reliability:** As this coding exercise was completed by the consultant who developed the MERL Guide, there is still a need in future for testing of inter-rater reliability of the methodology for all the operational indicators. There is an exception of the two Spanish projects in the sample which were reviewed and coded by a junior GPSA staff member with the oversight and guidance of the main coder (because the main coder does not speak Spanish). While the coder provided quality assurance and reviewed different versions, it was not possible for them to check all the details with 100% accuracy without being able to also review the source secondary Spanish documents. However, the first project in the sample was completed and reviewed by another GPSA team member who also reviewed and helped develop the MERL Guide and the GPSA Results Framework itself. This provided a degree of quality assurance and inter-coder reliability. However, this will continue to be tested as the GPSA continues to use the Results Framework and different grant partners and MERL Consultants use the required operational indicators and follow the MERL Guide for their assessment.

**Comparisons over time** It is also important to note that as this coding exercise was done retroactively, it cannot be used in a time series comparison over time. The data in future exercises will be both primary and secondary and will use the method as intended in the MERL guide. As such, the project and portfolio level values may be quite different as a result. Therefore, any comparisons drawn between this initial retroactive exercise and future ones must take into account the differences in method and available evidence.

**Concessions made for data gaps:** The coder used their own logical judgement for some aspects of assessing the indicators based on what was available, triangulation where possible and their knowledge of the GPSA conceptual framework for the operational indicators and intended results related to these indicators. In some cases, the coder erred in favour of positive performance as 'absence of evidence is not evidence of absence'. This has led to a higher aggregated values than may have been the case had more robust and detailed data been provided across all projects and if primary data collection had been possible. That said, there are details sources, quotes and page numbers referenced for all projects and indicators in the excel sheet and limitations are transparently explained.

## 2.6 Utility and validity

Despite the limitations and trade-offs noted above, there are several points to note that validate the usefulness and justify the validity of the 15 individual projects and aggregated sample dataset for the intended purposes of this retroactive coding and the SRP/Forward Look as follows:

- The testing with only secondary data available even with some documentation and quality limitations reflects the Results Framework indicators can be assessed using the MERL Guide methodology, even retroactively. Therefore, it provides a useful starting point for sense checking the indicators, methodology and the performance of previous grants against them.
- For determining to the coded values for every operational indicator of each of the 15 sample projects, the coder triangulated across several sources throughout project life and included numerous examples and quotes with page numbers (see individual project files). The inclusion of independent evaluation and ICRs in addition to self-reported grant partner documents provided a balance of different perspectives and a means to address bias (although the quality and detail of evaluations and ICRs vary which is important to consider). As mentioned above, the GPSA Secretariat's exchanges and tacit knowledge from working hands-on with grant partners is also a useful source of evidence to draw upon in future exercises for triangulation and to expand the evidence base.
- As the GPSA continues (including in the current round 4), the ability to conduct MERL on the operational indicators is expected to improve and become more reliable since there will be the same (functional equivalent) indicators across all grants. Once the indicators are more standardized, they can be measured in similar ways according to the MERL Guide. At the same time, grant partner project MERL system design, primary and secondary data will be aligned to enable more accurate, robust and ongoing MERL of these operational indicators. However, this also depends on a well-resourced MERL budget and recruiting MERL consultants throughout the project life that are of a sufficient quality and bring the requisite understanding and experience in the social accountability field.
- This process also demonstrates that it is unnecessary for GPSA projects to include more than a base set of key indicators to meaningfully assess and track results, performance and learning against the GPSA Results Framework and Theory of Action. Several of the projects in the sample had many indicators (20+) which stretches limited MERL resources and places more burden on the project grant partners than required.
- This process is also expected to become more routine and easier as grant partners, their MERL consultants and the GPSA staff become familiar with the Results Framework and MERL Guide over time with use of the aligned indicators and methodology, provided that sufficient MERL resources are available at the project and portfolio level.
- This exercise also indicated where there could be some adjustment or editing of the MERL Guide indicators methods in the future as it will continue to be iterated over time. It is suggested to keep the indicators and their methods as is for the current round of GPSA projects since it has already started, and considerable effort and time went into the MERL Guide and updated Results Framework. More testing of the current version is required and then updates can be well-justified and made. However, this document provides some reflections on changes that could be considered in future iterations.

Call for Proposals Round	Included in the sample	Excluded from the Sample	Justification for exclusion
Round One	Concerned Citizens Philippines	CARE Bangladesh	The excluded 4 projects do not have the min. docs of both ICR/final eval and minimal other reports that would enable a feasible or meaningful coding.
	Concern Mozambique	CARE Malawi	
	Expert Grup Moldova	DPI Kyrgyzstan	
	MJF SEBA Bangladesh	Economic Justice Network Malawi	
	Oxfam Dominican Republic		
	Oxfam Tajikistan		
	Wahani Visi Indonesia		
Round Two	Africa Freedom of Information Center Uganda		The two excluded projects had either zero (Moldova) or almost zero (Tunisia) project documents on file and therefore not possible to review.
	CARE Maroc	Center for Health Policies Moldova	
	Centro de Informacion y Recourses para el Desarrollo Paraguay	General Labour Union Tunisia (this gran was closed early)	
	Cordaid DRC		
	Globe International Center Mongolia		
	SAHA Madagascar		
Round Three	Save the Children Georgia	Akatiga Indonesia	All of the excluded projects do not have an ICR/final evaluation. Atatiga and YTSDE are still relatively early in their implementation and have therefore not done enough reporting or other MERL documentation yet.
	SEND Ghana	Eco-Dev Mauritania	
		IBIS Sierra Leone	
		Search for Common Ground Guinea	
		Transparency International Rwanda	
		World Vision Dominican Republic	
		YTSDE Indonesia	
Round Four	No projects included	PWYP Indonesia	Round four projects are not to be included because at the time of the coding exercise they were too new
		Anti-Corruption Action Center Ukraine	

		CARE Benin	in their start-up/implementation and have no or very limited reporting and results to assess yet vis-à-vis the GPSA Results Framework.
		INTEC-IHO Caribbean	
		VIVA Rio Brazil	
		East Europe Foundation Moldova	
		MASAM II Mongolia	

Table 3: Sample of GPSA projects for the retroactive coding exercise

### 3. Overview of Findings

The below table and section provide the aggregated values for all the operational indicators (and the output 3 GPSA/World bank non-operational indicator) that were coded for each project in the sample. The source excel file includes more details on the aggregation while it also includes the coded project files for all 15 sample projects with detailed explanations for each indicator and scores, with limitations and a reference to the sources referenced with quotes and pages numbers. It is important to reference the qualitative information that is available in these files in addition to the percentage and numerical values to interpret a nuanced and comprehensive understanding of the process and findings.

Table 4: Aggregated results for retroactive coding exercise

Outcome/output Indicator	Aggregated Value	Other relevant units of analysis or considerations of note
<b>Outcome indicator 1a)</b> Percentage of grant partners and their implementing CSO partners with improved capacity to engage meaningfully and collaboratively with one another and other stakeholders in the compact(s).	80% of projects (or 12/15) in the sample met this criteria for improved capacity of the lead grant partner and implementing CSO partners.	On the GPSA MERL guide scale of capacity improvement (1-4) the projects were rated as follows: 12 projects were rated as a '3' (Improved) 3 projects = were rated as a '2' (neither improved nor decreased) None of the projects received a '1' (decreased) or a '4' (significantly improved)
<b>Outcome indicator 1b)</b> Percentage of grants in which civil society actors, citizens and public sector institutions demonstrate improved capacity to engage meaningfully and collaboratively.	100% (15/15 projects)	N/A - yes/no indicator
<b>Outcome indicator 1c)</b> Percentage of grants that contribute to corrective measures taken by public sector actors to address proximate causes of service delivery failure.	100% (15/15 projects)	Yes/No was the main unit of analysis to code this indicator; however, where possible the percentage of corrective measures the project contributed to addressing was assessed with the available evidence.



		<p>A total of 7 of the sample projects yielded a percentage for corrective measures based on available data as follows:</p> <p>Oxfam Tajikistan = 85%  SEND Ghana = 61.5%  Save Georgia = 51%  AFIC Uganda = 36%  CCAG Philippines = 75%  SAHA Madagascar = 76%  CIRD Paraguay = 80%</p>
Outcome indicator 2) Percentage of GPSA grants in which civil society partnerships and relevant, targeted public sector counterparts engage in collaborative social accountability processes that include citizens.	100% (15/15 projects)	N/A – yes/no indicator
<b>Outcome indicator 3)</b> Percentage of GPSA grants in which collaborative social accountability processes targeted the proximate cause of service delivery failure.	100% (15/15 projects)	N/A - yes/no indicator
<p><b>Outcome Indicator 4a)</b> Percentage of GPSA grants in which governments seek to:</p> <p>(v) Use substantive lessons for improvements of targeted policies, processes, and mechanisms;</p> <p>(vi) Sustain elements of collaborative social accountability processes after the life of the project;</p> <p>(vii) Adapt insights from GPSA projects to scale them through programs or policies; or</p> <p>(viii) Apply elements of collaborative social accountability processes in additional localities or sectors.</p> <p>Note: this can be done through the government’s own reform program, donor-funded programs, or Bank-financed programs. [Target: 25%]</p>	100% (15/15 projects)	<p>Sustainability scores for each project also assessed on a scale from 1-5 (0%, 25%, 50%, 75% and 100%) – see full explanation under outcome 4 below.</p> <p>Breakdown of sample by sustainability ranking:</p> <p>1 (0%) = zero projects</p> <p>2 (25%) = 2 projects or 13.3% (CARE Morocco; SEBA Bangladesh)</p> <p>3 (50%) = 4 projects or 26.7% (SEND Ghana; Cordaid DRC; Save Georgia; CIRD Paraguay)</p> <p>4 (75%) = 4 projects or 26.7% (Wahani Visi Indonesia; Expert Grup Moldova; CCAG Philippines; Oxfam Dominican Republic)</p> <p>5 (100%) = 5 projects or 33.3% (Oxfam Tajikistan; AFIC Uganda; Concern Mozambique; TAME Mongolia; SAHA Madagascar)</p>
Output Indicator 1a) Percentage of GPSA grants with partner-led compacts	100% (15/15 projects)	N/A - yes/no indicator

with involvement from at least 3 fit-for-purpose stakeholder groups		
Output Indicator 2a) Percentage of GPSA of grants in which lessons learned during implementation informed course corrections.	87% (13/15 projects)	<p>Yes/No was the unit of analysis. However, the number of course corrections were identified and documented from the available evidence base for each project as below. For the 2 projects with only 1 identified course correction, they were counted as not meeting the indicator criteria:</p> <p>AFIC Uganda (9 course corrections)  CARE Morocco (12)  CCAG Philippines (8)  Concern Mozambique (7)  Cordaid DRC (5)  Expert Grup Moldova (7)  Oxfam Dominican Republic (1)  Oxfam Tajikistan (4)  CIRD Paraguay (1)  SAHA Madagascar (6)  Save the Children Georgia (7)  SEBA Bangladesh (5)  SEND Ghana (6)  TAME Mongolia (9)</p>
Output Indicator 3) Number of grants for which World Bank sector teams supported engagement between civil society and government	15/15 projects (or 100%).	<p>Yes/No criteria.</p> <p>It should be noted that 2 of the projects (Wahani Visi Indonesia &amp; Cordaid DRC) received a yes for the indicator criteria but had poor evidence from the World Bank side (i.e., incomplete or unavailable ICRs), with all sources from partner.</p> <p>Additionally, if a score of 'partial' were part of this indicator coding options, then 6 projects should have received partial based on the evidence which indicated the support was provided but needed improvement and was not a full 'yes':</p> <ul style="list-style-type: none"> <li>• AFIC Uganda</li> <li>• Concern Mozambique</li> <li>• Oxfam Dominican Republic</li> <li>• Save Georgia</li> <li>• SEBA Bangladesh</li> <li>• SEND Ghana</li> </ul> <p><i>Please note that there are a range of circumstances in each of these projects that impacted the level and type of support the World Bank could provide and also depended on the CSO partners and their preferences and engagement with the World Bank.</i></p>

As per the table and aggregated values above, one can draw the following key observations:

- All the projects in the sample had enough available documentary evidence to code the targeted indicators for this exercise, despite the limitations noted in section 2.

- The performance overall against the GPSA operational indicators is relatively high in fulfilling the criteria for the operational indicators as per the GPSA's Collaborative Social Accountability approach, concepts and processes and its aims for the portfolio, as reflected in the Theory of Action. This performance must be understood with the MERL guide and specific GPSA's concepts and understanding of 'what success looks like' for the GPSA. It also must be interpreted in line with the limitations and trade-offs made for this retroactive coding exercise.
- There are some additional units of analysis and data points that are helpful and important to include when interpreting the indicators such as for grant partner capacity (outcome 1a indicator); corrective measures (1c outcome); sustainability outcome 4); and lessons/course corrections (output 2 indicator).

The following provides a brief overview of key points of note for coding and aggregating each indicator and limitations in regard to the aggregated values.

### **Outcome 1a indicator**

*Overview:* A vast majority of the projects the sample were coded as fulfilling this criteria at 12/15 projects or 80%.

The GPSA has identified and prioritized the following specific five capacities<sup>1</sup> for Collaborative Social Accountability as follows:

- Civic capacity (2 types): intra-civic & inter-civic
- Organizational and operational capacity
- Analytical capacity
- Adaptive capacity

The scale defined by the GPSA (and included in the MERL Guide) for ranking each capacity is as follows:

- Deteriorated (1)
- Neither improved nor deteriorated (2)
- Improved (3)
- Significantly improved (4)

Additionally, the target setting, ranking and weighting for GPSA partner capacity assessments is further defined in the MERL Guide for future capacity assessments which will be done on all projects routinely as part of the MERL system for each project.

For each project, the coder used the available documentary evidence to determine the value of each project for this indicator. Specific examples and quotes are provided in each of the project coding files with sources and page numbers.

A yes was assigned to all projects that were above a 2. The percentage of total projects that met this indicator criteria was calculated by dividing the total number of projects that received above a 2 by the total number of projects in the sample - in this case 12/15.

*Limitations:* Based on the available evidence reviewed, the projects in the sample did not include formal capacity assessments in line with the GPSA capacities or based on the 1-4 scale to rank them. Therefore,

---

<sup>1</sup> See: Poli, Maria, and Maria F. Guertzovich. 2020. "Capacity and Implementation Support Area: Portfolio Performance Review." Global Partnership for Social Accountability Note 15. World Bank, Washington, DC.

the values provided on the scale for this indicator for each project is based on the coder's review of key documents and specific references to capacity development across the trajectory of the project.

In some cases, the GPSA capacities for Collaborative Social Accountability were referenced specifically, in other cases, it was more general. Therefore, the values provided on the scale must be interpreted with this limitation in mind.

### **Outcome 1b indicator**

*Overview:* All of the projects the sample were coded as fulfilling this indicator criteria. Therefore, the aggregated value is 100%.

A value of yes or no was the only unit of analysis for this indicator (as per the MERL Guide method). Therefore, there is no capacity assessment ranking as part of this indicator.

The coder used the available documentary evidence for each project to find examples for capacity improvement of different key stakeholder groups engaged in the CSA of the project: citizens, CSOs and public institutions across the trajectory of the project. Specific examples and quotes are provided in each of the project coding files with sources and page numbers.

*Limitations:* It is important to consider that the available documentary evidence for projects did not generally go into the level of detail necessary to determine with precision the capacity developed for each stakeholder across every project site. Nor did they use all the GPSA specific capacities. However, in many cases there were references to these capacities and the coder made the judgement to assign a yes based on projects having:

- Sample examples from several of the source documents that spanned across the different stakeholder groups, and.
- Evidence that capacity development was being done throughout the trajectory of the project.

Where available, the midterm and final evaluations were also key evidence for capacity development as independent sources. Many of the evaluations reviewed noted capacity development of these stakeholders as key results of the projects in the sample.

### **Outcome 1c indicator**

*Overview:* All of the projects the sample were coded as fulfilling this indicator criteria. Therefore, the aggregated value is 100%.

A value of yes or no was the only unit of analysis possible to code for this indicator for all the projects in the sample based on the available documentary evidence. In some cases, it was possible to provide the corrective measures % as the documents provided enough detail to do so.

The coder used the available documentary evidence for each project to find examples of specific corrective measures taken by public institutions based on the project's CSA processes and direct feedback to public institutions for improvement of service delivery in the given sector(s) and locations across the trajectory of the project. Specific examples and quotes are provided in each of the project coding files with sources and page numbers.

As all projects met the criteria across the full sample, an aggregate value of 100% was provided. In cases where a percentage of corrective measures could be assessed based on the evidence, this was also provided. See individual project pages for more details on how these values were determined.

*Limitations:* For the projects that were assigned specific % for corrective measures, this was based on percentages provided in the documentary evidence. In some cases, the coder needed to gauge and make an assessment. For example, if there were multiple percentages or differing ways of calculating them across the documents in a project compared to the GPSA MERL Guide method that should be used in future. While some final evaluations provided percentages and were a more reliable source, some are self reported by the grant partners. It was not possible to verify how percentages were calculated and the coder did not have access to raw data. Therefore, the percentages provided should be taken as approximate and not exact.

### **Outcome 2 indicator**

*Overview:* As all projects met the criteria across the full sample, an aggregate value of 100% was provided.

This was a Yes/No unit of analysis; however, it sought to assess if the targeting took place across all project sites. In 1 case, the review indicated that evidence was not available to determine this in detail (CIRD Paraguay). However, this does not mean that in all other projects, this outcome indicator was completely fulfilled across all the sites (nor would the GPSA expectation be for 100% achievement).

The coder used the available documentary evidence for each project to find examples for the project making concrete efforts and taking specific measures to analyse the political economy, stakeholders, context etc. on an ongoing basis and make decisions on targeting of proximate causes of service delivery failure based on this analysis, including changing course where necessary (linked to output 2 below). It also aimed to determine if this took place across the different sites of implementation. Specific examples and quotes are provided in each of the project coding files with sources and page numbers.

*Limitations:* Based on the secondary data documents available for the sample, it is difficult to ascertain with precision whether the Collaborative Social Accountability processes took place in all the project sites for all the projects. Also, the quality of CSA processes varies across the sites in terms of Collaborative Social Accountability engagement and effectiveness, as noted in many of the project source documents. However, it is not the GPSA expectation that that quality and engagement across all sites in a project would be the same and it is unrealistic to expect 100% achievement. The data available in future assessments would ideally demonstrate the differences. However, as a value needed to be given for this indicator, the coder determined that there was enough evidence in the documents to verify that there were at least CSA processes taking place in all projects with inclusion of citizens with several examples across the document sample for each project, therefore a yes was deemed as warranted for all the projects in the sample.

### **Outcome 3 Indicator**

*Overview:* As all projects were coded as fulfilling the criteria across the full sample, an aggregate value of 100% was provided.

This was a Yes/No unit of analysis; however, it sought to assess if the targeting took place across all project sites. In 1 case, the review indicated that evidence was not available to determine this in detail (CIRD Paraguay). However, this does not mean that in all other projects, this outcome indicator was completely fulfilled across all the sites.

The coder used the available documentary evidence for each project to find examples for the project making concrete efforts and taking specific measures to analyse the political economy, stakeholders, context etc. on an ongoing basis and make decisions on targeting of proximate causes of service delivery failure based on this analysis, including changing course where necessary (linked to output 2 below). It

also aimed to determine if this took place across the different sites of implementation. Specific examples and quotes are provided in each of the project coding files with sources and page numbers.

*Limitations:* Based on the secondary data documents available for the sample, it is difficult to ascertain with precision whether the requisite analysis and actions for targeting of proximate causes of service deliver failure took place in all the project sites for all the projects. Also, the quality of CSA processes varies across the sites in terms of CSA engagement and effectiveness, as noted in many of the project source documents. Similar to outcome 2 indicator, the GPSA does not expect that the quality and achievement to be the same across all sites, this would be unrealistic. However, as a value needed to be given for this indicator, the coder determined that there was enough evidence in the documents to verify that targeting for proximate causes took place in all projects, with several examples across the document sample for each project, therefore a yes was deemed as warranted for all the projects in the sample.

### **Outcome 4a) indicator**

*Overview:* All projects were deemed to meet at least a score of 2 or 25% sustainability (none received a 1 or 0%) therefore, an aggregate value of 100% was provided. However, the aggregate values for the breakdown of projects that fell into each category of 2-5 was also assessed and is provided. According to the GPSA target for 25% for outcome indicator 4a on sustainability, this has been reached and exceeded based on the retroactive coding exercise.

This indicator is fundamental to the GPSA model, as outlined in the Theory of Action. The GPSA's concept of sustainability as identified in the MERL Guide is very specific to CSA and the GPSA. It is also an experimental indicator that requires ongoing testing. It is also linked to other work and research ongoing by those involved with the GPSA working in the Transparency, accountability and Participation (TPA) sector to facilitate continuous learning about sustainability and scale-up of CSA.<sup>2</sup>

This outcome and indicator focus on the likelihood of key actors (public sector, WB, donors, INGOs) applying or sustaining elements and/or lessons of Collaborative Social Accountability (CSA) process. The GPSA is not attempting to achieve nor measure wholesale replication or full institutionalization of CSA. The emphasis is on process sustainability and function, not form of CSA. It casts a wide net on different potential outcomes for sustainability which are used by the coder to assess progress and evidence of sustainability for GPSA projects such as:

Work of the multi-stakeholder compact was integrated into another public sector project, program, policy.

- Public sector counterparts used lessons to inform public sector reforms
- Emulation by local public sector or service providers (e.g., schools) that observed, adopted or adapted the CSA.
- WB or other donors used lessons/CSA approaches to advise public sector/other partners' future programs
- WB or other donors will be funding an adaptation of the project moving forward in same sector or other sectors.
- Any observed or reported uptake, scale-up and sustainability led by other INGOs/CSOs.
- If the project trajectory demonstrates ongoing dialogue with key actors (relevant public sector officials and WB operations staff) over time, to move the process for potential CSA uptake forward.

---

<sup>2</sup> For more background and context on the GPSA's CSA sustainability approach and pathways to scale, please see <https://thomasmtaston.medium.com/introducing-a-resonance-pathway-to-scale-6cacd5163cd8>  
<https://medium.com/@florcig/how-context-shapes-pathways-to-scale-in-social-accountability-post-4-of-5-d417cfe2b4f5>  
<https://medium.com/@florcig/scale-up-in-time-revisiting-how-we-evidence-process-context-6c53f82a1817>

However, sustainability criteria are not fulfilled just through information sharing/dissemination. Efforts to engage and find avenues for sustainability by project grant partners also must take place throughout the project, not just at the end.

The ranking of sustainability for the GPSA projects is defined by the scale and criteria in the table below:

*Table 5: Sustainability criteria and ranking rubric as per GPSA MERL Guide*

Score	Description
1	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.  The unit of measurement for this indicator in the Results Framework is %. Therefore, a score of 0% would be provided for the indicator in the Results Framework and considered as 'no uptake'.
2	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.  In this instance, a score of 25% would be provided for the indicator in the Results Framework.
3	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).  In this instance, a score of 50% would be provided for the indicator in the Results Framework
4	Evidence of dialogue with priority stakeholders and/or public sector institutions on how to adopt, adapt and/or sustain elements of the CSA process in future operations, policies, or programs.  In this instance, a score of 75% would be provided for the indicator in the Results Framework
5	Evidence of actions taken by priority stakeholders and/or public sector institutions on adoption, 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.  In this instance, a score of 100% would be provided for the indicator in the Results Framework

For each project, the coder used the available documentary evidence to find examples for the project making concrete efforts and taking specific measures to engage key stakeholders, programmes, policies etc. to find entry points, dialogue and opportunities for sustainability in different forms an ongoing basis (not just at the end of the project, but from the onset). Specific examples of sustainability in line with the GPSA criteria were also identified and where possible triangulated with both the reporting and the independent evaluations. Specific examples and quotes are provided in each of the project coding files with sources and page numbers.

*Limitations:* It is important to note that this indicator and the assessed scores provided represent a projection of the likelihood of public sector institutions applying or sustaining lessons from or elements of the social accountability process after the project life. This is based on signs of sustainability and uptake, drawn from the evidence collected each year; emphasis is not limited to the final evaluation. The “seek

to” is a key part of this indicator statement; the uptake of GPSA projects is contingent, in that it can be introduced but not sustained by the project after the end date (alternatively the opposite can be true with sustainability/scale-up results taking place post-project; however it requires follow-up to verify and track these developments which is often unfeasible with GPSA final evaluation/ICR timelines and available resources). It is unrealistic for the allocated time, resources, and budget invested in GPSA projects to expect and results in wholesale uptake and institutionalization; success for GPSA in terms of sustainability can be many things.

The GPSA designed the project/portfolio Results Framework and MERL guide e.g., indicators, to a focus on function and processes do not form. The number and ambition of indicators risks creating incentives for the project team to invest in activities to checkboxes and reach targets, rather than invest efforts in an effective way. Furthermore, it is also the aim of the GPSA to foster adaptive learning across the projects and it is therefore important to leave room for flexibility and changes to CSA approaches that are adapted to a given context, based on MERL evidence, rather than standardizing forms of CSA which may not be suitable.

The assessment scale for this outcome indicator in the MERL Guide is experimental and being tested (as with all the criteria) and as there is such a wide range of potential options for sustainability to be met, detouring the sustainability scores for the sample projects included many different examples that are not the same (i.e., a score of 50% on one project may look quite different in form than a score of 50% on another).

As the documentary evidence available did not assign numerical or percentage scores to the projects in terms of sustainability because this was not a requirement at the time of designing these projects, the coder made the assessments for this indicator based on their judgement; the defined GPSA criteria; and the relevant information available in the project documentary evidence. Therefore, the qualitative explanations and examples that accompany the sustainability indicator for each project are important to read alongside the score for the nuanced details.

Overall, each project was deemed to meet 2 or more with examples provided across the source documents, and in many cases, the final independent evaluations.

### **Output 1 Indicator:**

*Overview:* As all projects were coding as fulfilling the criteria across the full sample, an aggregate value of 100% was provided. Even if in 2 cases the evidence specified that it was less than 100% of sites, the overall value was still high enough to warrant a 'yes'.

This was a Yes/No unit of analysis; however, the coder sought to assess if the CSA compacts with at least 3 fit-for-purpose stakeholders meeting regularly took place across all project sites where the evidence was available to determine this.

The following projects had evidence to indicate compacts did not achieve the full indicator across all compacts:

- Cordaid DRC 67%
- SAHA Madagascar 93%

Similar to other indicators, the GPSA expectation is not that the projects will achieve 100% success for compact quality and engagement across all target sites. The projects that are able to provide disaggregated results are reflective of the level of data that the GPSA seeks to yield from its projects. However, as a value needed to be assigned for each project, the coding gave 100% to all sample projects except for those two above, as the percentage was explicitly stated in these two projects' documents.



However, this does not mean that all the other sample projects necessarily fulfilled this criteria across all compacts/sites; it means that the documents did not provide this level of detail and therefore 100% was assumed for the purposes of this coding exercise as it would be unfeasible to determine another value. For the purposes of this coding and due to the limitations in the evidence available, 13 of the projects were assessed as 100% of the compacts/sites. It is very likely though that quality and levels of engagement varied across sites/compacts for each project as would be expected given different contexts and challenges and dynamics at the compact/site level.

The coder used the available documentary evidence for each project to find examples for the project targeting and engaging fit-for-purpose stakeholders (relevant to the context, sector and country) on an ongoing basis in the CSA process for decision making, consultation, collaboration etc. It also aimed to determine if this took place across the different sites of implementation generally and the quality although this is a limitation overall (see next column). Specific examples and quotes are provided in each of the project coding files with sources and page numbers.

*Limitations:* The evidence base overall provides ample examples of multi-stakeholder consultations and meetings including several fit-for-purpose stakeholder groups. However, the documents for the sample do not generally provide this level of detail at the site/compact level nor the of frequency of meetings. Therefore the 100% score provided for this RF indicator for GPSA strategic review and aggregation purposes must be interpreted with the limitation that the score does not account for or confirm that a specific minimum frequency for each compact and site of implementation was met. Nor does it provide qualitative assessment of the differences in terms of stakeholder engagement between each compact/site. However, the frequency of compact members seems to be regular based on the documentary evidence reviewed; therefore, this indicator can be considered as achieved for all projects.

### **Output 2a) Indicator:**

*Overview:* Several relevant examples of learning for course corrections were identified in 13 of the 15 projects, while 2 of the 15 only had 1 and therefore this could not be considered as ongoing practice. Therefore, a score of 13/15 was calculated to come to the aggregate value of 87% (rounded from 86.7%)

For each project in the sample, the coder reviewed all documents to source examples of relevant lessons that led to intentional course corrections/adaptations to the project (model, operations, strategy, etc.). The total number was calculated with the qualitative information and source for each one provided in the individual project coded sheets. The examples range but they are based on analysis, MEL, context shifts, etc. It was also important to demonstrate that this was taking place by the grant partner/project on an ongoing basis as a regular practice in order for the indicator criteria to be deemed as achieved. Therefore, a Yes/No is provided along with the number of course corrections identified for each project. All course corrections with sources and page numbers are included in each project file.

*Limitations:* The documentary evidence provided is a sample of the overall evidence base for all projects. All the lessons and course corrections/adaptations to the project may not be reflected in the documentation overall. Therefore, it is not guaranteed that the number of course corrections provided to each project is exact or comprehensive. That said, the majority of projects had several examples noted in the documents which means the criteria can be fulfilled as a yes. For this retroactive coding exercise (and future assessments), it is less important that the number is precise as long as the evidence demonstrates that a project made intentional course correction based on lessons learned from MEL, analysis and context factors throughout the project for improvement. The causal analysis provided in the qualitative narrative accompanying the scores are essential to providing meaning and rigour to the assessment.

It is also relevant to note that the 2 projects that were assessed as not meeting this indicator criteria were done by a different coder than the rest of the projects as they were in Spanish and had to be sourced to

another GPSA team member who was junior and did not develop nor was deeply familiar with the MERL guide criteria. Although they received guidance and review from the main coder, since this indicator requires reading of all documents in Spanish to identify all the specific examples of course corrections, the main coder could not quality check this indicator in the way they could for the other indicators. This means that there is a chance there was a different interpretation to this indicator and if it were done by the same coder, the score may be different.

### **Output 3 Indicator:**

*Overview:* It is important to note that this indicator ((Number of grants for which World Bank sector teams supported engagement between civil society and government), is not an operational one and therefore was not included in the MERL Guide, at the time of the coding exercise, a method had not yet been fully developed. The GPSA asked the consultant to include it based on the available evidence, and it was agreed that a simple yes or no would be used to determine the value for the purposes of this exercise.

As all projects were coded as fulfilling the criteria across the full sample, an aggregate value of 15 projects or 100% was provided. This indicators the World Bank's support for meaningful engagement between civil society and government (output 3 indicator) has been met in all projects at least to some degree with many projects yielding several examples of sector, TTL and GPSA support for the grant partner projects in the form of capacity building, trouble shooting, brokering linkages between the project and key stakeholders, and providing hands-on guidance to projects.

This is a Yes/No indicator therefore the coder determined the value by looking across the evidence base for specific examples of how the WB country staff, TTL and the GPSA supported the project in different ways such as brokering relationships and linkages with key stakeholders, supporting problem solving and barriers, providing guidance and capacity building, conducting monitoring visits, etc. The coder also ensured that this was evident across the trajectory of the project for the indicator to be deemed as met. Particular evidence that was useful for this indicator were the ICR reports as they assess WB performance on the project, the final evaluations as they are independent, and the progress report sections in which the grant partner provides specific feedback on the GPSA and TTL support. Specific examples and quotes are provided in each of the project coding files with sources and page numbers.

*Limitations:* In interpreting the aggregate values the following limitations and biases should be noted:

- Several of the ICR reports were not completed in full and lacked the necessary information on TTL/WB support or the ranking that is meant to be included.
- The reporting on WB support in progress reports by the grant partner could be influenced by power dynamics of the donor/grant recipient nature.
- Not all the evaluations available in the project sample included reference or assessment to WB support; where this was available, the score can be deemed as more substantiated due to its independence.

A 'partial' score for some of the projects would have been more appropriate as there was evidence of support but also evidence of areas for improvement, missed opportunities or insufficient support. However, given the Yes/No choice, all projects in the sample had enough evidence to warrant a yes rather than no.

To understand which projects achieved a full 'yes' and which ones would have received a 'partial' if that was an available indicator assessment value, it is important to reference the specific project coded files and the narrative information for this indicator.

The coding process therefore revealed that this indicator and coding frame would benefit from a score of 'partial' or converting it to a ranking scale rubric system (1-3 or 1-5 approach) and none of them

warranted a no, based on the evidence available, however there are several projects that would not be deemed a full 'yes' if partial options were available in the coding frame.

#### 4. Observations and recommendations

The following table provides some observations that the consultant coder documented through undertaking this exercise regarding the process, information management, data access and quality. Where relevant or possible, there are recommendations included that could be considered for the GPSA in general related to relevant data/evidence for projects/the GPSA and future Results Framework MERL exercises.

Table 6: Key observations and recommendations

Observations and recommendations based on this retroactive coding exercise		
MERL area	Observation	Recommendation
Access to evidence for coding exercise	The process for the consultant to receive clear and accurate information on the number/overview of all the GPSA projects, the project documentation available and the copies of the documentation took a few months. Several files are duplicate copies or drafts, with some documents incomplete or of poor quality. This reflects that the information management and storage of key GPSA project documents and oversight of data quality is perhaps weaker and more arduous that it could and should be.	Information management and data quality assurance should be strengthened for GPSA projects, especially if the coding and aggregation of GPSA performance against its Results Framework is to take place on a regular basis in order to improve the efficiency of these exercises.
Baselines	Many baselines focus on the external (such as a survey of water users) but not information that is useful and relevant to assessing and analysis for the operational Results Framework indicators. For instance, there is a lack of analysis and data on capacity (outcome 1) (although not using the MERL Guide and GPSA capacities nor formal assessments and ranking); existence/experience with Collaborative Social Accountability (outcome 2; the proximate causes of service delivery failure (outcome 3); or identifying potential actors & entry points for uptake (outcome 4) and info that is highly relevant for adaptive learning/ course correction already.	Ensuring that baseline reports include data collection and analysis for all the required operational indicators in future GPSA project baselines would support a more comprehensive evidence base for MERL of the Results Framework and increase the feasibility for comparing the context and values from the project start through to the end.
Mid-term reviews/ evaluations	Many of the projects lacked mid-term reviews or evaluations in the documentary evidence base reviewed which means was difficult to triangulate progress and performance of the Results Framework indicators for the project sample with an independent source.	Where possible, and in line with the MERL Guide, mid-term reviews/evaluations are recommended. These can be lighter touch than final evaluations but should include the assessment of the operational Results Framework indicators.

<p>Final Evaluations</p>	<p>Similar to baselines, many evaluations reviewed did not explicitly reference or provide analysis/findings on the GPSA Theory of Action and required operational Results Framework indicators. While this is partly due to the operational indicators not being in all projects reviewed, the evaluations did not even refer to those Results Framework indicators that were included in a given project. This meant that the coding relied more on the values from the ICR reports which were not substantiated and well-explained. There are exceptions and the coding was still able to glean useful information from the evaluations to support the coding, however it seems relevant that evaluations should include an explicit assessment of a projects' Results Framework indicators and the underpinning logic and trajectory of the GPSA Theory of Action.</p>	<p>It is important for final evaluations to explicitly link to GPSA Theory of Action as a key objective of assessment and learning while they should also include the indicators from the project Results Framework directly and ensure all the domains are assessed in addition to any other objectives or themes of an evaluation.</p> <p>The GPSA's requirement for all projects to include the functional equivalent of the operational indicators in the GPSA Results Framework, the MERL research design process, and the standard KEQs (as per the MERL Guide) is expected support this process and assurance in future. In addition, updating the standard GPSA Key Evaluation Questions that are in the MERL guide and Consultant ToR would also help to ensure that the necessary components of the projects are assessed and enable easier portfolio comparisons and aggregation of results.</p>
<p>Consistency of MERL consultants</p>	<p>Consistency in using the same MERL consultant over time indicates positive outcomes for the quality and rigour of the MERL for the projects reviewed and supports the ability to code the GPSA's operational Results Framework indicators. It should be noted also that the quality of the consultant reports differs but with the strongest ones, there is value add to the projects. This often implies a higher level of resources that need to be dedicated to the consultants for MERL of GPSA projects in order to yield high quality MERL for a project and to enable the consistent and robust coding of GPSA Results Framework operational indicators at the project level and to aggregate them at the portfolio level.</p>	<p>The 2014 practice of using the same MERL Consultant (as per the GPSA's model MERL Consultant Terms of Reference and the MERL Guide) is supported by the findings of this retroactive coding exercise. However, with the MERL budgets available for GPSA projects, it may be difficult to resource the required long-term and quality MERL consultant support to enable this – the GPSA should re-examine its MERL budgets and consider increasing them where possible.</p>
<p>Capacity development and assessment</p>	<p>The evidence base overall for the project sample reviewed did not include capacity assessments in line with the identified GPSA capacities for Collaborative Social Accountability. While some of the projects may have conducted some form of capacity assessments, this was not part of the documents reviewed or available. Many documents reviewed in the evidence sample</p>	<p>In order to glean the level of detail necessary to assess projects (lead grant partners and their direct implementing CSO partners) across identified types of GPSA capacities, project evaluations and reports should explicitly specify the types of capacities and speak to each one more distinctly. The MERL Guide</p>

	<p>make ample reference to ‘capacity’ improvements and some note those relevant to the GPSA capacities. However, this was not provided with ample specific details, which made assessing the core identified GPSA capacity types challenging. The coder relied on a broad conceptual framework for capacity improvement relying on the available examples in order to conduct the exercise and provide the scores for outcome indicator 1a). However, this would be limitation to coding and scoring of the indicators on capacity development if it is not addressed going forward.</p>	<p>includes specific guidance and suggested method for coding to be conducted by the project’s MERL Consultant which is expected to improve this in future.</p>
<p>Corrective measures</p>	<p>The coding in terms of the % of corrective measures (PDO level indicator for all projects or outcome indicator 1c) was difficult to assess with precision in the dataset although there were exceptions. However, these were often presented in ICRs and evaluations in a way that was not possible to validate and in mixed/varied ways even within a given project. The language ‘corrective measures’ was not always used.</p>	<p>In order to score accurately for PDO indicator under outcome 1c) it will be important for projects and the MERL Consultant to ensure that all corrective measures are tracked, documented and compiled throughout the project life by the grant partner in a way that can be more easily counted, assessed and validated to ensure the integrity and validity of the indicator value in addition to confirming that the project contributed to corrective measures (yes or no). It is important to try to ensure they use the language <b>corrective measures</b> so it is clear in the grant partner reporting and evaluations and can be easily identified for Results Framework analysis/coding by the MERL Consultant and substantive in independent evaluations.</p>
<p>Components</p>	<p>Components referenced in in some projects’ core documentation are different to the standard GPSA components 1-3 (also reflecting the GPSA Theory of Action) that are meant to be in the project paper, therefore it could be confusing to relate to components in this way if they are different than the GPSA’s.</p>	<p>The TTL/GPSA should ensure that there is consistency of these components and prevention of confusion in the reference and documentation to project components in Project Papers (or equivalent) and other project documentation, including evaluations/ISRs/ICRs,</p>
<p>Compacts</p>	<p>The output indicator 1 is concerned with the number of fit-for-purpose stakeholders that meet regularly (compact) across the target sites of project implementation, including the national level. While the project documentation generally included several references to key stakeholder engagement which provided the basis for the coding in this exercise, it lacked the level of detail to assess this indicator to the</p>	<p>For the depth of analysis and coding of output 1 and to enable overall portfolio aggregation, each project should include a list of each compact at each target level/site with members and meetings. Ensuring that the final list included in key grant partner documentation, especially in the final evaluation</p>

	intended degree of the MERL Guide and GPSA Results Framework indicator and to enable accurate aggregation.	and ICRs Results Framework even if as an annex.
Narrative details in ICR Results Frameworks	In several cases, the qualitative information that is meant to accompany the Results Framework scores was lacking and highly inconsistent across the sample dataset of the Results Frameworks of project ICRs. Stating just a yes/no or 'achieved' without the narrative explanation is a barrier to accurate coding and aggregation.	The TTL/GPSA should ensure that the Results Framework indicators scores each year and used for the ICR (provided by the MERL Consultant/grant partner) are all accompanied by a comprehensive narrative description with relevant qualitative details before completing/approving a project ICR.
ICRs	Upon the review of ICRs in the project sample, there were some that were incomplete and therefore did not provide the information that would be necessary and expected to support the project coding of the operational Results Framework indicators.	The ICRs are a key piece of documentation for GPSA projects and the assessment of performance against the operational Results Framework indicators at the project level and aggregation of the GPSA Results Framework at portfolio level. More attention and assurance for comprehensive and complete ICRs is strongly recommended. For a 'best practice' example of an ICR reviewed in the sample, please see TAME Mongolia and CCAG Philippines
Output indicator 3	As this indicator ((Number of grants for which World Bank sector teams supported engagement between civil society and government), is not an operational indicator and therefore not included in the MERL Guide, at the time of the coding exercise, a method had not yet been fully developed. The GPSA asked the consultant to include it based on the available evidence, and it was agreed that a simple yes or no would be used to determine the value for the purposes of this exercise. However, in the process of coding and scoring this indicator it was observed that simply proccing a 'yes' or 'no' is too crude and lacks nuance – as a result coding in this way will end up being mostly 'yes' and poorly reflect the reality of mixed/varied support in a given project.	As explained in section 3 above, it is suggested that this indicator coding frame and method be adjusted in future to be on a scale or allow for a 'partial' score for this indicator.  For independent verification and to account for the asymmetrical power dynamics between grant partner and WB, The GPSA should also consider that projects include a Key Evaluation Question on this indicator for the MERL consultant to independently assess or that it be part of the Results Framework indicators that they are expected to include in the annual project Results Framework assessments/updates in addition to the operational indicators.
Grant partner Results Framework Indicators	Many of the grant partner project Results Framework include far too many indicators and some were also placed at the wrong level for PDO or Immediate Results (IR) (sometimes IR	This is expected to improve by having each of the GGPSA's operational Results Framework included in the grant partner's

	<p>ones should be PDO ones as they are about overall impact and results). This places undue burden and stretches limited MERL resources for the project.</p>	<p>Result Framework (localized to the project with functional equivalents) with the support of a qualified MERL Consultant throughout the project life. While grant partners may choose to include additional indicators, it is also recommended that the grant partners avoid adding too many as it is not required or necessary for the GPSA and uses limited MERL resources.</p>
--	--	---