GPE 2022–2026 Study: Country-Level and Thematic Evaluation

Final Inception Report

Prepared by Triple Line, Learn More and Technopolis

June 2023

Document Control

Document title	GPE 2022–2026 Study: Country-Level and Thematic Evaluation: Final Inception Report
Reference no.	WOR-2500J
Prepared by	RAA, JC, MK, AK, LN, AP, ES, GZ
Checked by	CMP
Date	29 June 2023

Contents

List of abbreviations

2LT	Learning Leadership Team (of the GPE Secretariat)
CEP	Country Engagement and Policy (team of the GPE Secretariat)
CSO	Civil Society Organization
CTL	Country Team Lead
EMIS	Education Management Information System
EOL	Education Out Loud
EPL	Education Policy and Learning (team of the GPE Secretariat)
FGO	Finance and Grant Operations (team of the GPE Secretariat)
GEA	Girls' Education Accelerator
GPE	Global Partnership for Education
GPE 2020	GPE's Strategic Plan for 2016–2020
GPE 2025	GPE's Strategic Plan for 2021–2025
ITAP	Independent Technical Advisory Panel
ITRP	Independent Technical Review Panel (for the evaluation)
KIX	Knowledge and Innovation Exchange
MEL	Monitoring, Evaluation and Learning
MLT	Multiplier Grant
MSC	Most Significant Change
PEA	Political Economy Analysis
PILC	Performance, Impact and Learning Committee
QA	Quality Assurance
R&P	Results and Performance (team of the GPE Secretariat)
RISE	Research in Improving Systems of Education
SCG	Systems Capacity Grant
SDG	Sustainable Development Goal
SEAH	Sexual Exploitation, Abuse and Harassment
SMART	Specific, Measurable, Achievable, Relevant, and Time-bound
SNA	Social Network Analysis
STG	System Transformation Grant

UIS	UNESCO Institute for Statistics
UNESCO	United Nations Educational, Scientific and Cultural Organization

1 Introduction

This inception report sets out the evaluation approach and methodology for the evaluation of the Global Partnership for Education's (GPE's) global and country-level support to partner countries under the GPE strategic plan for 2021–2025 (GPE 2025). A consortium, led by Triple Line with partners Learn More and Technopolis, has been engaged to conduct this four-year evaluation (2022–2026).

The purpose of this inception report is to outline our overall approach to the evaluation, including its conceptual and evaluation framework and detailed methodological approach. We will apply a phased approach to our evaluation, allowing us to adapt and evolve our methodology in successive phases. Therefore, while this inception report will provide an overview of the approach used for all three phases of the evaluation, the approach and methodology will be revisited at the end of each phase.

1.1 Process to Develop the Inception Report

In order to refine the approach and methodology described in this inception report, the evaluation team undertook a number of activities during the inception phase (November 2022 to April 2023). These activities included the following:

- On-boarding sessions facilitated by the GPE Results and Performance (R&P) team, with additional sessions with a select number of other teams and individuals across the GPE Secretariat, including from the Finance and Grant Operations (FGO) Team, the Country Engagement and Policy (CEP) Team and the Learning Leadership Team (2LT).
- Comprehensive desk review, including GPE strategic and programmatic documents (shared by the GPE Secretariat) and literature review on system transformation in education.
- Consultation sessions held across the GPE Secretariat (including with the core management team and across the CEP team) and with a selection of constituencies of the GPE Board.
- A co-design and theory of change workshop, hosted by the R&P team at the GPE Secretariat offices in Washington DC in February 2023.

A summary of our consultation sessions (and feedback received) and the co-design and theory of change workshop can be found in **Annex 1**. These activities allowed the evaluation team to better understand the design and roll-out of the GPE operational model and to tailor our evaluation design.

1.2 Organization of the Report

The inception report is divided into the following sections:

- **Section 2** describes the context for the evaluation, including background on GPE and our understanding of the GPE 2025 operating model and its implementation.
- **Section 3** lays out the purpose and scope of the evaluation and introduces the evaluation's overarching approach, guiding principles and limitations. More information on our sampling for our country-level case studies is found in **Annex 3** and the research protocols which follow on from our guiding principles are found in **Annex 8**.
- **Section 4** provides the framework for our evaluation, including an explanation of our conceptualization of system transformation, the implications for the theory of change for the evaluation and the linkages to our evaluation matrix. This section is accompanied by annexes including the full description of the assumptions of our theory of change.
- **Section 5** outlines the methodology we will use throughout the three phases of this evaluation. This includes an overview of the methodological approach and how we will adapt it to our country-level case studies, then outlines the steps involved in the desk research, primary data collection, analysis and reporting. This section also includes a discussion on the identified limitations of our methodological approach. Our research protocols are provided in **Annex 8**. Our suggested outline for the country-level case study and synthesis reports can be found in **Annex 4**.**Annex 3**
- **Section 6** explains our learning approach and provides an indication of potential learning products (contingent on further discussion with R&P).

- **Section 7** summarizes our evaluation workplan and key deliverables.
- **Section 8** outlines our evaluation management strategy, including information on our evaluation governance and reporting, communications and stakeholder engagement, risk management, quality assurance, data management and standards, and team organization. We have further annexed additional information on our communications protocols for the evaluation (**Annex 5**), our risk management approach (**Annex 6**) and risk matrix (**Annex 7**), quality assurance framework (**Annex 9**) and team organogram and roles (**Annex 10**).

2 The Global Partnership for Education

2.1 Background on GPE

GPE is a multi-stakeholder partnership and is the world's largest global fund dedicated to transforming education in lower-income and lower-middle-income countries. Working with over 85 partner countries, GPE mobilizes partnerships and investments to accelerate access, enhance learning outcomes and promote gender equality through equitable, inclusive and resilient education systems fit for the 21st century.

GPE was established in 2002 (as the Fast Track Initiative) to bring together partners from lower-income countries, donors, international organizations, civil society (including youth and teacher organizations), the private sector and foundations to transform education systems to support the world's most vulnerable children, including girls, children with disabilities and those living in fragile and conflict-affected states, many of whom are displaced.[1]

GPE targets educational changes at the systems level, with a focus on strengthening education systems and leveraging domestic and international finance to accelerate and sustain progress to deliver at least one year of preschool and 12 years of quality education for every girl and boy.[2]

At the heart of the partnership are governments and development partners. At the global level, GPE is governed by a Board of Directors, which includes representatives from each partner constituency including partner countries, donors and representatives from civil society, the private sector, private foundations and multilateral agencies. The Board sets the Partnership's strategy, debates policy and allocates funds. Operations and administration of GPE are supported by the GPE Secretariat, who facilitate coordination, convening and collaboration across the partnership. At the country-level, the partnership comprises local education groups[3], who support government-led coordination to implement planned transformations to improve the education system. Furthermore, at the country level, GPE mechanisms are supported by a Coordinating Agency, which is selected by the local education group to support and facilitate local education group work, who in turn support the selection of a Grant Agent to support partner countries in the preparation and implementation of GPE grants.

2.2 The GPE 2025 Strategy and Operating Model

The GPE Strategic Plan for 2021–2025 (GPE 2025) aims to accelerate access, learning outcomes and gender equality through equitable, inclusive and resilient education systems fit for the 21st century, in alignment with Sustainable Development Goal (SDG) 4.

The strategy brings together GPE's mission to mobilize partnerships and investments that transform education systems in developing countries, leaving no one behind. Responding to demand from GPE partner countries, GPE 2025 seeks to bring change, building on previous successes but finding new ways to accelerate progress toward education systems that are both more resilient as well as founded on equity, inclusion, quality and efficiency.

GPE 2025 focuses on eight priority areas, in which it seeks to have transformative impact: access; learning; gender equality; inclusion; early learning; quality teaching; volume, equity and efficiency of domestic financing; and strong organizational capacity.

---

[1] GPE website, "About GPE".
[2] GPE Factsheet, "Global Partnership for Education" (2023).
[3] Local education groups typically comprise representatives from the ministry of education, local government, line ministries and ministry of finance, as well as from development partners including civil society organizations, teacher groups and representatives, international organizations and the private sector. The local education groups also include the Grant Agent and the Coordinating Agency. More information on the role of local education groups can be found at: https://www.globalpartnership.org/node/document/download?file=document/file/2019-10-GPE-principles-effective-local-education-groups.pdf

To achieve its goals, GPE has developed a new operating model, learning lessons from the findings of evaluations on the previous strategic plan[4] and from consultations with stakeholders including partner country representatives. These findings and consultations pointed to a need for three strategic shifts in GPE's ways of working:

- Strengthen mutual accountability for system transformation, including the alignment and harmonization of external financing and to coalesce country-level actors' resources and capabilities around country-owned system transformation priorities.

- Sharpen GPE funding for system transformation, including an expansion and systematization of funding to support country capacity development, funding to support the unlocking of system transformation through catalyzing progress in factors that enable system transformation and embedding monitoring, evaluation and learning (MEL) through all these processes.

- Mobilize strategic capabilities to reinforce country capacity for system transformation, through more efforts to support advocacy and collaboration.[5]

The new operating model is also guided by GPE's principles of effective partnership, which builds on GPE's Charter and acknowledges the need to increase decentralized mutual accountability, drive national government ownership and strengthen its capacity, rebalance the country-level model to focus on implementation and sector policy dialogue, and reduce GPE processes and transaction costs. The GPE 2025 operating model responds to these priorities and articulates an approach to system transformation (figure 1). We elaborate further on our understanding of system transformation in the context of GPE in section 4.1.

ACHIEVING QUALITY EDUCATION FOR EVERY CHILD

Emphasis on gender equality

ACT, LEARN & ADAPT
- ABILITY TO LEARN FROM EVIDENCE AND ADAPT AS NEEDED
- 9 Implementation of policy reform and monitoring of partnership compact
- 8 Additional partner support
- 7 Selection of GPE grant agent – GPE grant application

ASSESS & DIAGNOSE
- 1 Dialogue within local education group
- 2 Identification of key policy reform area
- 3 Analysis of enabling factors
- RECOMMENDATION BY ITAP AND COUNTRY ENDORSEMENT

PRIORITIZE & ALIGN
- FULL ALIGNMENT OF SYSTEM TRANSFORMATION PRIORITIES
- 6 Board review of allocation amount
- 4 Evidence on key policy reform area incorporated
- 5 Creation of partnership compact

GPE MOBILIZES PARTNERS TO SUPPORT SYSTEM TRANSFORMATION

SUPPORTED BY
- Knowledge & Innovation Exchange (KIX)
- Education Out Loud
- Strategic capabilities

Figure 1. Summary of GPE's approach to system transformation as part of GPE 2025[6]

---

[4] This includes the Independent Summative Evaluation of GPE 2020 (conducted during 2019–2020) and a series of 28 summative and prospective country-level evaluations which concluded in 2020.
[5] GPE Meeting of the Board of Directors Document BOD/2020/11/12 DOC 05 "GPE 2025 Operating Model Framework."
[6] GPE Factsheet, "GPE 2025: Operationalizing a system transformation approach" (2022).

The basis of the GPE 2025 operating model is the development of a country-owned **partnership compact** by local education groups[7] which supports mutual accountability and government ownership. The process for developing a partnership compact encourages[8] partner countries to **assess and diagnose** a key policy reform area through joint policy dialogue and using an analysis of four key enabling factors: data and evidence; gender-responsive sector planning, policy and monitoring; sector coordination; and the volume, equity and efficiency of domestic public expenditure on education. GPE provides support to this process through the Secretariat, and a GPE-convened Independent Technical Advisory Panel (ITAP) reviews the enabling factors analysis and provides independent feedback to country-level stakeholders and the GPE Board on the most important challenges with respect to the enabling factors. Partnership compacts are approved at the country-level. The GPE Board does not play a role in approving partnership compacts, but uses ITAP assessments to support decisions on GPE funding.

The priority reform is articulated in the partnership compact itself and includes the roles and responsibilities of different country-level stakeholders including any remaining gaps, allowing country-level partners to **prioritize and align** their efforts around a common transformative goal.

Learning is embedded into the partnership compacts in order to ensure that partner countries are able to **act, learn and adapt** through the use of data and evidence to inform effective and efficient sector policies and programs. Learning is further supported by the GPE Secretariat, which provides additional technical support, monitoring and completion frameworks with minimum standards for GPE grants and the GPE Evaluation Policy guides the use of evaluation across GPE's support.[9]

To help partner countries advance their education systems through the operating model, GPE provides both **financial support** (in the form of grants) and **non-financial support** (through efforts to drive policy reform, promote partnerships, support knowledge and evidence, and advocacy work).

The financial support mechanisms used as part of the GPE 2025 operating model include:

- **System capacity grant (SCG):** $1 million to $5 million over a five-year period. The SCGs provide flexible funding to support capacity development needs of partner countries as categorized as high priority areas during the initial assess and diagnose phase of the GPE 2025 model. The grant targets the development of capacities to enable country-level partners to conduct analysis, gender-responsive planning and data collection to support the compact development process, as well as longer-term education system planning, coordination, financing, learning from evidence and adapting implementation.

- **System transformation grant (STG):** variable funding of $5 million to $162.5 million determined through an allocation formula (including a minimum allocation portion and a top-up portion) targeted toward supporting low-income countries, vulnerable lower-middle-income countries and small island and landlocked developing states. The funding allows partner countries to implement programs that contribute to system transformation.

- **Multiplier grant (MLT):** $5 million to $50 million matching fund facility available to all GPE-eligible countries. Supports country-level partners to mobilize additional resources by leveraging external funding from partners such as multilateral development banks, bilateral donors, non-traditional funders, private sector and foundations, alongside GPE funds.

- **Girls' Education Accelerator grant (GEA):** up to $25 million additional funding to support 30 eligible partner countries, who have the largest needs in relation to girls' education. Access to the GEA is not a standalone grant, but linked to either a MLT or STG (or both).

In addition to financial support, GPE provides further support to system transformation at the country-level through the following:

---

[7] At the time of writing of the Inception Report (April/May 2023), GPE is in the process of assessing modifications and adjustments to the operating model as part of ongoing learning (further described in section 2.3).
[8] This is a requirement for all system transformation grant-eligible countries and optional for non-system transformation grant-eligible countries.
[9] GPE Factsheet, "GPE 2025: Operationalizing a system transformation approach" (2022) and information available on the GPE website "How to apply for grants."

- **Access to strategic capabilities:** GPE supports this process through collaboration with global and regional partners to provide additional capabilities to support priority reforms.

- **Knowledge and Innovation Exchange (KIX):** GPE provides partner countries with access to innovations and expertise by providing funding to South-based research and knowledge exchange mechanisms between partner countries to fill gaps in education policy. KIX aims to provide evidence-based solutions to national policy makers to support policy dialogue and planning processes. It is managed by the International Development Research Centre.

- **Education Out Loud (EOL):** EOL is GPE's fund for advocacy and social accountability. It aims to strengthen civil society advocacy and government accountability. It supports civil society to be active and influential in shaping education policy that meets the needs of communities and vulnerable and marginalized populations. EOL is managed by Oxfam IBIS.[10]

Gender is hardwired through the process of country-led system transformation. More information on how GPE hardwires gender throughout GPE 2025 can be found in Box 1.

Box 1. GPE 2025: Gender equality at the heart of education system transformation[11]

Gender equality is at the center of new operating model and key to enabling education system transformation. Under GPE 2025, GPE has a new, ambitious approach to promote gender equality to, within and through education.

In order to hardwire gender throughout the operating model, GPE does the following:

- Supports comprehensive country dialogue on gender equality by requiring countries to examine gender-responsive planning and monitoring as one of the enabling factors for analysis in the compact development process, ensuring that the inclusive dialogue at country-level includes organizational and/or individual gender expertise and requires that gender equality considerations are included in the priority reforms identified in partnership compacts.

- Demonstrates their own financial commitment to gender equality by putting gender equality at the center of its grant mechanisms. The programs funded by the SCG, STG and MLT must include activities to address gender-related barriers to a quality education faced by girls and boys and the GEA provides an additional incentive to support countries to address barriers to girls' education.

- Puts gender equality in and through education and girls' education on the world stage through global advocacy efforts and through EOL and sharing knowledge and innovation on gender through KIX. Gender equality has also been set as one of the key knowledge priorities as part of KIX's extension in 2023.

2.3 Roll-out of the GPE 2025 Operating Model

In December 2020, the GPE Board approved the framework for the GPE 2025 operating model and its phased roll-out with the pilot (e.g., first cohort) starting January 2021.[12] Table 1 outlines the cohorts of partner countries and the planned cohort start date.

Table 1. GPE 2025 operating model's phased roll-out, by cohort (as April 2023)

Cohort #	Countries	Cohort start date
Cohort 1 (Pilot) (6)	Democratic Republic of the Congo, El Salvador, Kenya, Nepal, Tajikistan, Uganda	January 2021
Cohort 2 (14)	Bhutan, Burkina Faso, Cambodia, Comoros, Fiji, Gambia The, Senegal, Sierra Leone, Somalia – Somaliland, Somalia – Puntland,	October 2021

---

[10] GPE 2025 Strategic Plan.
[11] More information on GPE's approach to gender equality can be found on GPE's website (What we do – gender equality"), GPE Factsheet "Girls' Education: The Path to Progress" (2022) and GPE's blogpost on "Hardwiring Gender Equality in GPE 2025" (2022).
[12] GPE Meeting of the Board of Directors Document BOD/2020/11/12 DOC 05 "GPE 2025 Operating Model Framework"

	Somalia – Federal, Tanzania – Mainland, Tanzania – Zanzibar, Zimbabwe	
Cohort 3 (16)	Benin, Chad, Côte d'Ivoire, Guyana, Kiribati, Maldives, Marshall Islands, Federated States of Micronesia, Nicaragua, Niger, Rwanda, Samoa, Sao Tome and Principe, Solomon Islands, Tonga, Tuvalu, Vanuatu	January 2022
Cohort 4 (10)	Ethiopia, Cabo Verde, Djibouti, Eritrea, Guinea, Lesotho, Liberia, Madagascar, Nicaragua, South Sudan	July 2022
Cohort 5 (31)	Afghanistan, Bangladesh, Burundi, Cameroon, Central African Republic, Democratic Republic of the Congo, Dominica, Ghana, Grenada, Guinea-Bissau, Haiti, Lao People's Democratic Republic, Malawi, Mali, Mauritania, Mozambique, Myanmar, Nigeria, Pakistan - Balochistan, Pakistan – Khyber Paktoonkhwa, Pakistan – Punjab, Pakistan – Sindh, Papua New Guinea, St. Lucia, St. Vincent and the Grenadines, Sudan, Syrian Arab Republic, Timor-Leste, Togo, Republic of Yemen, Zambia	January 2023

As of April 2023, all five cohorts have begun roll-out. Figure 2 demonstrates the progress thus far of each cohort against key stages of the operating model.
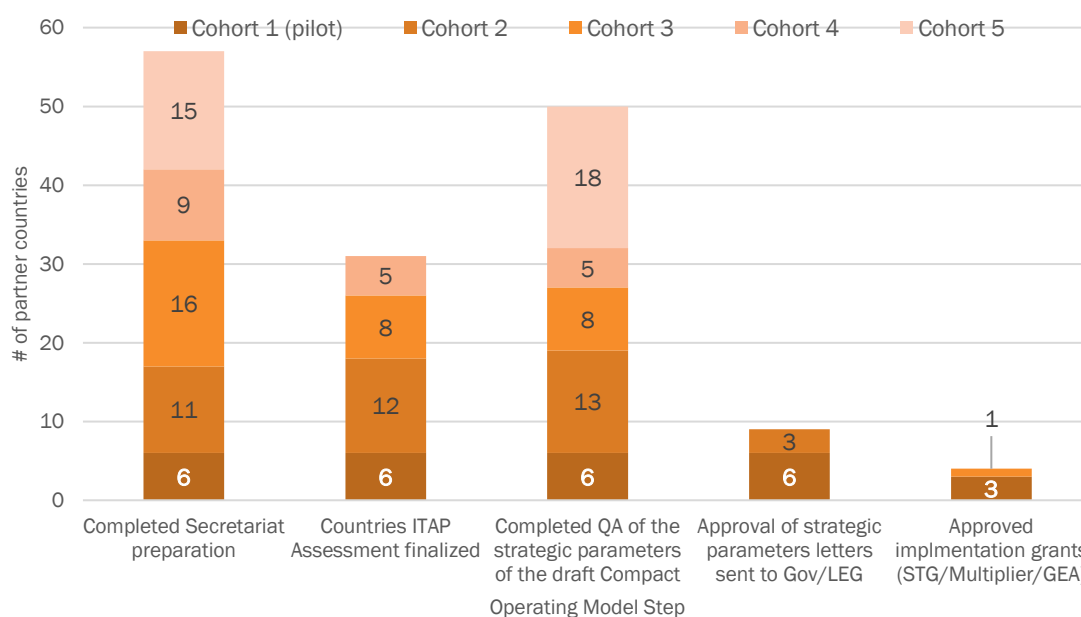
Figure 2. Progress of the roll-out of the GPE 2025 operating model, by cohort as of April 2023[13]

Alongside the roll-out of the model, the GPE Secretariat has developed technical guidelines to support partner countries, held a series of partnership engagement webinars to understand partner information needs and to provide tailored training and dialogue, as well as developed a learning function to capture the experiences of the pilot and record lessons learned to identify potential adjustments for the operating model.[14]

To support learning from the pilot, the GPE Secretariat conducted data collection and held a two-day webinar in June 2022 to hear reflections from stakeholders in partner countries who participated in the

---

[13] Data from GPE dataset, "New OM Pipeline As of April 19 2023."
[14] GPE Meeting of the Board of Directors Document BOD/2021/12 DOC 04 "Operating Model Roll-Out" Annex A.

pilot. The reflections[15] revealed both promising changes initiated as a result of the operating model as well as ongoing areas for improvement:

- **Country dialogue:** the operating model process has supported a more consultative country dialogue process as a result of the enabling factors and compact development process. However, improvements are needed to ensure that the technical working groups responsible for leading the process are inclusive of civil society, that the political aspects of the reform are not de-emphasized, and that policy prioritization is not just informed by the availability of funding and political willingness to work in the priority reform.

- **Gender:** the pilot also demonstrated that the use of gender-responsive planning and monitoring as one of the enabling factors meant that this topic was included in the planning process and development of the partnership compact. There is still a need for the use of data and evidence on gender, including a gender analysis to inform prioritization and strengthen contextual analysis and attention to the intersectionality of gender with other forms of disadvantage barriers, as well as an approach to monitoring progress on gender equality.

- **Transaction costs:** pilot countries took longer to complete the compact development process, experience delays in the enabling factors self-assessment and ITAP review stages.

The ongoing learning was to have informed adjustments to the operating model that did not require Board decisions. Findings such as these have already informed important adaptations to the operating model process; through the learning framework and through this evaluation, GPE seeks to ensure that feedback and evidence can continue to inform both further adaptations as well as GPE's future strategic direction.

---

[15] Lessons learned and reflections were shared as part of GPE's 2022 Results Report (section 2.5) and the GPE blogpost "Learning from the Partner Countries piloting of the GPE 2025 Approach" (2022).

3   The Country-Level and Thematic Evaluation

3.1 Purpose and Scope of the Evaluation

This evaluation is a multi-year, phased study that evaluates and generates continuous learning on GPE's global and country-level support to partner countries as delivered through GPE's new operating model, and outlined in the GPE 2025 strategy. The evaluation will cover GPE's support starting from the beginning of the roll-out of the GPE operating model (from January 2021) until mid-2026.

The objectives of the evaluation follow:

- Understand the continuous relevance, coherence, efficiency and effectiveness of GPE's global and country-level support (including the operating model and strategy), vis-a-vis countries' system transformation efforts and with respect to GPE's key thematic areas.
- Identify promising practices and explain their underlying success factors in context.
- Formulate recommendations on how GPE's operating model and strategy could strengthen to support partner countries' transformative reform processes at country or global levels.

To fulfill these objectives, we will conduct a multi-year evaluation study that is collaborative, developmental and sensitive to complexity in order to support key evaluation stakeholders in efforts of ongoing learning. The evaluation will focus on country-level case studies which will be synthesized into an evaluation of overall GPE support through the operating model and overall strategy.

The primary audience of this evaluation is the GPE Board and the Performance, Impact and Learning Committee of the Board (PILC), GPE Secretariat, partner country and other GPE partners/stakeholders. At the Secretariat-level, our evaluation can help guide improved practices, frameworks, guidelines, approaches to financing and tools for country support. The evaluation can support the Board to consider GPE's strategy, governance mechanisms and processes for the allocation and use of financing. Finally, the evaluation will also seek to support partners at the country-level with evidence and learnings to support more effective engagement with GPE processes and inclusive, evidence-based and effective transformative reform for their education system.

3.2 Overarching Evaluation Approach

This evaluation focuses on the GPE 2025 strategy and operating model (including efforts such as technical engagement and advocacy at the global level, that seek to reinforce the operating model), how the operating model works to support GPE country partners to identify priority reforms, the suitability of the operating model for facilitating design and implementation of priority reforms and progress toward system transformation, and how the operating model works in alignment with countries' existing education policy systems and agendas as well as the wider education systems landscape. The evaluation will examine both GPE's financial and non-financial support to reform efforts.

We designed an evaluation that is theory-based, uses realist and developmental evaluation approaches and is phase-based, with gender hardwired throughout its methodology and processes. The evaluation will involve a set of country-level and thematic case studies conducted across the evaluations' three phases. This section sets out the different approaches that guide our evaluation. Further information on the evaluation questions that will be investigated through these approaches is found in our evaluation matrix (section 4.3) and details on how we will operationalize these approaches through data collection analysis and reporting is found in section 5.

3.2.1 Theory-Based, Realist and Developmental Evaluation Approach

Evaluating system transformation is challenging. It takes a long time for outcomes to emerge, or substantive systems change to occur. The pathways are rarely linear, and the path may change as new evidence emerges (in the spirit of learning and adaptation). This is why using a theory-based approach incorporating principles of realist and developmental evaluation are central to this evaluation.

Our evaluation is therefore guided by these three complementary evaluation approaches.

First, our evaluation is **theory-based** – it is underpinned by an overall theory of change, a tool which sets out the logic of GPE's approach and how it intends to achieve its objectives, starting with the country's engagement with the development of a compact and the use of the GPE 2025 operating model to support the process, through to the outcomes of the priority reform and how the operating model supports education system transformation. Basing the evaluation on a clear theory of change allows us to articulate the linkages (and embedded assumptions) between the intervention's inputs and activities, then to outputs and outcomes. The evaluation then seeks to test these linkages and assumptions using evidence collected as part of the evaluation in order to determine whether the theory implied in the GPE 2025 strategy leads to its intended changes.

We have developed a theory of change and a set of assumptions for the evaluation with the support of the GPE Secretariat through a co-creation workshop held in Washington DC in February 2023. The theory of change and assumptions are introduced in section 4.2 (assumptions are further detailed in **Annex 2**). Together, the theory of change and assumptions seek to articulate how the activities of GPE, through its financial and non-financial support mechanisms and the use of the GPE 2025 operating model processes, are intended to lead to the alignment of relevant stakeholders in identifying, designing and implementing transformative education reform at the country-level. We also use our theory of change to outline how the implementation of the selected transformative reform, alongside further country efforts and GPE Secretariat and other partner support, can lead to subsequent change for the education system.

We will use a **realist evaluation approach** to test and refine the evaluation's theory of change. The realist approach is based on the hypothesis that outcomes are determined by the context in which they occur. Therefore, a realist evaluation approach seeks to ask the question "what works, for whom, under what circumstances and how."[16] Translated to this evaluation, we will use a realist approach to reflect the diverse contexts of GPE partner countries interacting with the GPE model, by accounting for if and how context affects the achievement of intended outcomes, as articulated by our theory of change.

To support the realist approach, our evaluation will include an investigation of the political economy of education systems as part of our country-level case studies.[17] As well as political economy, the realist evaluation approach also recognizes the importance of the effect of other non-observable entities, such as the power dynamics of partnerships. We account for these entities and the broader context in the assumptions we have developed for the theory of change.

The evaluation is **developmental** in that it will generate ongoing and participatory learning for GPE to understand the extent to which the theory of change is playing out as expected and supporting system transformation. The focus is on supporting reflection, dialogue, learning and decision-making during the lifetime of the evaluation, rather than just delivering a point-in-time judgment with recommendations. In practice this means liaising with key stakeholders through the Secretariat, reporting emerging findings and embedding learning into the initiative.

Developmental evaluation principles complement both approaches by ensuring a strong emphasis on learning throughout the evaluation. The goal of this developmental approach is to provide regular feedback that improves roll-out in real time.[18] This is suited for programs that operate in multiple, complex or uncertain environments, where innovative approaches are used and where adaptations can be made to the intervention and these developments can be mapped. We believe these features hold true for the GPE 2025 strategy and operating model. Developmental evaluation also requires that systems thinking be incorporated into the evaluation design and thus particularly suits the evaluation of an intervention aiming for system transformation.

Throughout the evaluation, we will feed back findings to the GPE Secretariat, the GPE Board (particularly PILC) through both pre-defined and on-demand learning moments to support GPE in the

---

[16] The term "realist evaluation" was first used by Pawson and Tilley in their book "Realistic Evaluation" Sage Publication Ltd, 1997.

[17] Following changes in the political economy over time is also an important component of our phased evaluation approach.

[18] Developmental evaluation was introduced by Michael Quinn Patton in his 2010 text "Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use" (Guildford Press., 2010) as a choice which is also responsive to context and suitable to assess innovation, complex issues and crises, for example.

implementation of the operating model. The evaluation team will work closely with the GPE Secretariat to determine the most appropriate and relevant ways to conduct learning activities with country-level and global stakeholders, based on evaluation findings. We further describe our approach to learning and learning products in section 5.

3.2.2 Phase-Based Evaluation Approach

The evaluation will be implemented over four years (2023–2026) in order to follow the roll-out of the GPE operating model for 2021–2025. We employ a **phased-based approach** to our evaluation, which allows the evaluation to longitudinally track the roll-out of the GPE operating model. We use phases for both the following:

- Progressive focus of the evaluation – in other words, we will examine the progressive phases of a partner country's involvement in the roll-out of the operating model and the implementation of the associated priority reform). For each of these phases, the evaluation will have a different focus and set of evaluation questions.

- Implementation of our evaluation – for example, we will stagger the roll-out of our country-level case studies, in order to capture different phases and iterations of the operating model, as it itself is rolled out in partner countries staggered across cohorts.

Table 2 provides an overview of the progressive focus of the evaluation, and how the evaluation will focus on different questions to reflect the different stages of a partner country's engagement with the operating model. We use the terms baseline, midline and endline to help denote these different phases.

Table 2. Overview of the phased focus of the evaluation

	Baseline	Mid-term	Endline
Status of countries engagement with the GPE operating model	Design and align resources to transformative reform	Implement transformative reform	Progress toward transformative reform and associated impact
Progressive focus of our evaluation	To explore how countries design transformative education reforms and their capacity to implement the reform To evaluate the relevance and effectiveness of GPE support to identify and design reforms with potential for system transformation	To examine countries' progress in implementing the transformative reform To evaluate the effectiveness of the GPE operating model in supporting the early implementation of transformative reforms	To assess early signs of progress and results of the transformative reform To evaluate the effectiveness of the GPE operating model in supporting progress toward/achievement of system transformation

By revisiting a country over time, our evaluation can progressively assess the coherence, efficiency, effectiveness, and (potential for) impact and sustainability of the operating model to the identification, design and implementation of the country's chosen transformative reform. We will examine the country's readiness for and progress toward change, the effectiveness of past and present reforms (and potential differences between the two) and the pace at which change occurs. This includes aspects of both **process** and **summative** approaches to evaluation. To reflect this, our evaluation matrix will include progressive evaluation questions to track how change unfolds over a series of activities, GPE support and time, as visualized in our theory of change. We provide more information on the evaluation matrix in section 4.3.

We will also stagger the evaluation's roll-out through its country-level case studies, which will be conducted in two waves. This allows us to account for the phased roll-out of the operating model over a series of cohorts, allowing us to examine partner countries that have been part of the early roll-out of the operating model (cohorts 1 and 2) as well as those rolled out later (cohorts 3, 4 and 5). Depending on how and when changes to the operating model are made, this allows us to also examine changes in the operating model by sampling from countries which have experienced different iterations of the operating model.

Finally, our phased approach also allows us to learn and adapt our evaluation design and methods, to ensure that they are suited to how the operating model's roll-out is unfolding. We divided our evaluation into three time-bound, annual phases (Figure 3). This allows us to, at the end of each phase, summarize our findings, share learnings and reflect upon the results of the evaluation and make changes to our methodology if necessary. This allows us to devise or refine evaluation questions as we progress through the evaluation.

At the end of each phase, we will produce short country-level case study reports as well as a synthesis report that draws together findings from across the country-level case studies and findings on specific priority themes. Further details on reporting are found in section 5.6. At the end of Phases 1 and 2, we will also produce a short concept note which puts forward a more detailed plan for the next stage of the evaluation, as well as any proposed changes to the evaluation design or methods from the previous stage.

3.2.3   Approach to Country-Level Case Studies

To allow us to closely examine the GPE 2025 strategy and operating model and how it plays out in different contexts, we use a case study approach with a country as the unit of analysis. Our country-level case studies will be longitudinal studies in which, over time, we will explore the ways in which countries identify and undertake transformational education reforms and progressively assess the contribution of GPE's support and processes in helping countries to identify and implement these reforms.

In total, we will examine **15 countries**, which will reflect a variety of different contexts and geographies of GPE partner countries. As described in the previous section, we will stagger the start of our country-level case studies, allowing us to include partner countries participating in different cohorts of the operating model roll-out.

We will begin with eight country-level case studies in the first phase of the evaluation, undertaking the baseline for this first wave of country-level case studies, progressing to the midline in Phase 2 and endline in Phase 3. At the start of Phase 2, we will roll-out an additional seven country-level case studies in our second wave, completing their baselines in Phase 2 and then the midlines in Phase 3. Figure 3 demonstrates this phased approach to our country-level case studies.

	Phase 1 (Apr 2023 to Jan 2024)	Phase 2 (May 2024 to Jul 2025)	Phase 3 (Jul 2025 to Jun 2026)
Baseline	Wave 1 (eight countries)	Wave 2 (seven countries)	
Midline		Wave 1	Wave 2
Endline			Wave 1

Figure 3. Demonstration of the waves of country case studies across phases

For Wave 1 country-level case studies, we selected eight countries from the pool of those that had completed development of their partnership compact prior to the start of Phase 1 of the evaluation: **Cambodia, Democratic Republic of the Congo, El Salvador, Nepal, Tajikistan, Tanzania – Mainland,**

Sierra Leone and Uganda. For Wave 2, we will be similarly limited to selecting from those countries that have completed their compact development process prior to the start of Phase 2. Further details on how we selected these eight and how we will select countries for Wave 2 can be found in our sampling approach in Annex 3.

3.2.4 Approach to Thematic Case Studies

GPE's work as part of GPE 2025 cuts across eight priority, or thematic, areas: access; early learning; equity, efficiency and volume of domestic financing; gender equality; inclusion; learning; quality teaching; and strong organizational capacity. We will select themes to form thematic case studies to examine how GPE 2025 priority areas are identified, epitomized and adapted at the country-level and to what extent the GPE operating model empowers country partners to do this through the compact development process. The evaluation will focus on the themes of gender equality and domestic financing, as these two themes are the most embedded throughout the operating model.[19]

- **Domestic finance case study:** we will assess how partner countries identify and address bottlenecks with respect to domestic finance and how GPE's financial and non-financial support (e.g., cross-national mechanisms) contributes to this process and helps advocate for prioritization of domestic finance reforms.

  We will also look at how the policies and priority reforms chosen by the partner countries address domestic finance and whether the chosen reforms increase the likelihood of leading to necessary enabling conditions to transformative reforms. At the outcome level, we will examine the extent to which the operating model has incentivized addressing gaps in domestic financing and led to greater stakeholder alignment in this area. At the process level, we will study how the compact strengthened partner countries' ability to identify, diagnose and address domestic finance-related bottlenecks, and how the advocacy efforts led to greater visibility around domestic finance-related policy reforms.

- **Gender equality case study:** given the importance of gender equality to GPE as core to the operating model and as one of the key priority areas, gender equality is both a thematic focus for our evaluation as well as a principle embedded in our evaluation approach. Similar to the case study on domestic financing, our gender quality case study more closely examines the ways in which partner countries understand and conceptualize gender quality, have the capacity to identify and address system bottlenecks with respect to gender equality. We will also assess the ways in which partner countries have undertaken policy actions and prioritized reforms to support transformational gender equality and the contribution of GPE's financial and non-financial support for partner countries to do so.

For both case studies, we have formulated specific evaluation questions, which can be found in our evaluation matrix. The thematic case studies use data collected from the 15 country-level case studies. To account for progress in the two thematic areas both cross-sectionally and longitudinally, we will collect and examine data on gender equality and domestic financing at baseline, midline and endline of each country-level case study.[20]

We will also conduct cross-country analysis on the other six GPE priority areas (access, learning, inclusion, early learning, quality teaching and strong organizational capacity), which will be presented in annual synthesis reports. The basis of this analysis will once again be data collected from our country-level case studies. This analysis will explore both the emerging intended and unintended outcomes in each of the priority areas as well as whether and how the education sector reforms in partner countries target these six GPE priority areas. However, the data to support analysis in these remaining six priority areas will be limited by the extent to which the partner countries in our sample are targeting reforms and transformational outcomes in these priority areas.

---

[19] The evaluation's ability to examine all the priority areas is limited by the extent to which partner countries identify priority reforms which fall under each thematic area.
[20] We will discuss with GPE in planning for Phase 2 whether mid-line case studies should inform a thematic analysis of gender only.

3.2.5 Approach to Gender Hardwiring in the Evaluation

As outlined in section 2.2, gender equality is at the center of the new operating model and key to enabling education system transformation. It is embedded in the first strategic objective of GPE 2025 and is one of GPE 2025's priority areas. As per GPE's Evaluation Policy, the evaluation will assess how gender has been hardwired throughout the GPE new operating model and which gender equality outcomes have been reached in countries as a result of GPE's gender approach (or which gender quality outcomes to which GPE support has contributed).

Moreover, we will adopt a gender-sensitive approach in all evaluation activities, by analyzing disaggregated data, ensuring voices of women and minorities are heard and promoting gender balance and non-discrimination within the evaluation team. Throughout data collection, the perspective of different stakeholders, especially in accessing education, will be incorporated into our sample and collected by a carefully balanced team of women and men of different cultures and ethnicities.

At the outcome level, we will examine the effectiveness of the operating model to promote changes in gender equality, exploring intended and unintended outcomes in the field of gender equality to, within and through education. Findings will be systematized and analyzed in the corresponding thematic case study. At the process level, we will examine the way in which gender has been hardwired throughout the work of GPE from the Secretariat and Gender Hub to the local education groups and country levels.

This approach is reflected in the evaluation matrix, where gender considerations are mainstreamed across all levels and criteria. The gender analysis principles we will rely on are described further in section 5.5.1

3.3 Guiding Principles and Research Ethics

In carrying out this evaluation, we will adhere to GPE's Evaluation Policy, which emphasizes independence and impartiality, credibility, transparency, utilization for learning, ethical principles and participation.[21] We represent these principles in the following ways:

- **Independence and impartiality:** our evaluation team has full independence from the GPE Secretariat. We have no vested interest in any particular area of the evaluation or in the findings. Each of our team members has been vetted for any conflict of interest, as will any future team members. While the ultimate accountability of the evaluation falls with our evaluation Project Director, the evaluation is led by a core team comprising members from all three consortium members and one member external to all three organizations, in order to mitigate against any bias. We work closely with GPE's R&P team to facilitate full access to GPE reporting information and allow for our full autonomy in carrying out data collection, analysis and reporting.

- **Credibility:** to ensure the credibility of our evaluation, we have assembled a team that brings together expertise in a number of critical aspects for the evaluation, including complex evaluation methodologies and education system transformation. Our inception phase has included extensive consultations across GPE to ensure that a sound understanding of the GPE 2025 strategy and operating model informs our evaluation approach and methodologies. The credibility of the evaluation will be further tested through a review of the evaluation's design by key evaluation stakeholders such as the GPE Secretariat's R&P team and the Independent Technical Review Panel (ITRP) advisors assigned to this evaluation, and subsequent revisions and updates by the evaluation team.

- **Transparency:** we will uphold the principle of transparency by working with R&P to ensure that aspects of our evaluation design and process are publicly available, including this inception report, and a detailed description of the evaluation process is completed, as part of the evaluation reports.

- **Utilization for learning:** through our developmental evaluation approach, our evaluation is participatory and utilization focused. We will work closely with the GPE Secretariat to co-design each step of this evaluation to ensure that the evaluation is both relevant and timed to support key

---

[21] In addition, GPE's Evaluation Policy also includes principles on gender equality and inclusion, which are described above in section 3.2, and on capacity development and joint evaluation. The latter two, we believe, are more internal-facing for the GPE Secretariat and therefore have not included them here.

decision-making moments both within the GPE Secretariat and for the GPE Board. We will also ensure that the evaluation activities are socialized at the country level in the countries in which evaluation activities will take place to support country-level stakeholders, such as members of the local education groups and notably the ministries of education, to engage with our evaluation findings. The Secretariat's regional clusters and country teams will be involved in this country-level utilization process.

- **Ethical principles:** our evaluation adheres to the highest ethical principles for evaluation. The consortium partners are value-driven organizations committed to strong protocols on ethics in their work. This includes ensuring that we build **trust and respect** into all engagements with both key evaluation stakeholders and evaluation participants throughout the evaluation process. We also apply the **principles of human rights** throughout data collection to ensure that we have considered a variety of aspects of inclusivity and that no harm is done to any participant engaged through the evaluation. We hold **high standards for safeguarding**, requiring all of our team to comply with Triple Line's Code of Conduct and Safeguarding Policies, and monitoring compliance and adherence. Finally, we apply **confidentiality and privacy** and a policy of **voluntary participation and informed consent** throughout our data collection and ensure that our evaluation participants understand this. More information detailing all of this is found in **Annex 8**.

- **Participation:** we will ensure that diverse and distinct views are included as part of our evaluation. The use of country-level stakeholder mapping, a political economy approach, a gender-sensitive approach and in-person fieldwork ensure that our evaluation considers the power dynamics embedded in education systems and offers both detailed thinking and flexibility to ensure that diverse views are included in the evaluation. For country-level studies, we will work with and through country-level partnership structures, such as local education groups, for consultation in the design of the country-level case studies as well as in disseminating findings.

3.4 Limitations to the Evaluation Approach

In this section, we note some limitations inherent to the overall approach chosen for this evaluation, as well as remediations where available. Methodological limitations are considered in section 5.7.

- This evaluation uses a case study approach (both country-level and thematic) to progressively examine the GPE 2025 strategy and operating model and its applications in different contexts. Our sample includes 15 country-level studies, a number which we believe will allow us to deeply explore a diversity of contexts and achieve a variety of rich case study findings. However, as we are only able to examine 15 out of a possible 77 countries in which the model will be rolled out – and there will doubtless be considerable variation in both the country context and in how the operating model is rolled out – there is a limit to the extent to which we will be able to generalize our findings across the population of countries eligible for participation.

- The GPE operating model is being rolled out using a cohort approach. This allows GPE to iterate on its operating model to make improvements across different cohorts. While our phased approach allows us to stagger the roll-out of country-level case studies to examine different iterations of the model, it should be noted that this may limit the ability to compare across the Waves 1 and 2 country-level case studies if significant adaptations to the model were to be made. However, these differences will allow us to examine whether the newer adaptations are more suitable for GPE's overall aim of system transformation and how they fare against findings from the earlier cohorts.

- An important criterion of country case study selection is the completion of the compact development process. An assumption as part of our sampling approach is that we anticipate that most (if not all) eligible countries will have completed their compact development process prior to the start of Phase 2, meaning that most will be eligible to participate in our sample based on the planned timeline. However, early observations from the roll-out of cohorts 1 and 2 suggest that there has been a wide range in the rate at which partner countries have been able to progress through the operating model process. If this remains true for successive cohorts, we may find that our sample becomes inadvertently biased by including a greater number of countries that display factors that support quick progress along the operating model process, and similarly include fewer (or exclude entirely) countries with slower than planned progress. As we select country cases for Wave 2, we will apply a stratified approach to ensure that the overall sample of 15 country-level

case studies covers a diversity of contexts and regions to balance against any potential biases from across the earliest operating model participants, and consider whether there are any observable patterns across those excluded from our sample due to delays in compact development and if we may need to revise our selection criteria accordingly.

- The current planned timelines for compact development for cohort 5 are such that there is a risk of countries from this cohort being ineligible for our Wave 2 country selection if they experience delays in completing their partnership compacts. If particular aspects or features of the operating model are only implemented for cohort 5 countries, we may have a more limited opportunity to explore these if our sample did not purposively include partner countries from cohort 5. Therefore, as part of the selection criteria for Wave 2 sampled countries, we will pay particularly close attention to the roll-out and availability of cohort 5 countries for inclusion in our sample.

- Given the time frame of this evaluation, even countries from across the first cohort of the operating model (and including our Wave 1 country-level case studies) will have had limited time (two years) in which to implement their priority reform and progress toward outcomes before our endline data collection. In addition, the endline questions at Phase 3 will likely be applicable only to the eight Wave 1 countries, based on the pace of roll-out of the model. As a result, while the evaluation will seek to identify early signs of progress and results of the transformative reform, we acknowledge that the time frames required for the fruition of observable and measurable evidence of transformative change at the level of the learner, and particularly change and results/impact that can be attributed back to GPE's operating model, are likely to be beyond the scope of this evaluation. We will consider the priority reform within the historical context of that particular sectoral area (which will be captured as part of each baseline) in order to help place the reform in a broader policy making continuum. Also, the use a forward-looking line of inquiry is intended to attenuate this limitation – we will use prospective questions to examine whether the right conditions are in place and risks, which when considered together, may suggest the likelihood of success of the reform.

- Theory-based evaluation is often used to support causal analysis by establishing and testing the causal link between observed (or expected to be observed) changes. However, while a theory of change is often used to support the determination of attribution or contribution, we note that the complexity and scope of the theory of change developed for this evaluation adds complications to this process. As the evaluation will examine the way in which the GPE operating model has or has not brought together key country-level stakeholders to identify a specified priority reform, we believe it is possible to attribute the relevance and effectiveness of the priority reform as transformational to GPE. However, it is not possible to solely attribute the effectiveness of the implementation of the priority reform to bring about systems-level changes and transformative outcomes at the level of learners, to the GPE operating model. Here, it is more appropriate to examine the contribution of GPE's financial and non-financial support, in conjunction with the factors outside of GPE's influence by working backward to trace and link observable outcomes to contribution claims. However, as discussed above, we do not expect that such outcomes will be evident within the time scope of this evaluation.

- The evaluation remains very ambitious in its intended scope and coverage as expressed in its evaluation questions. Furthermore, the developmental approach and its focus on usability will mean that the evaluation must produce short, focused and timely outputs, which to some extent will limit what can be covered. It will be critical after Phase 1 that the evaluation team appraise how effectively the evaluation can collect data (especially monitoring data on the implementation of the priority reform, and on the education results/impact achieved) and answer evaluation questions within the specified time frames of a phase in order to contribute to learning. Revisions to the evaluation design and approach may be required to sharpen or focus the evaluation and data collection, as well as to improve the effectiveness of its methodology.

4 Evaluation Framework

4.1 Conceptualizing System Transformation in Education

4.1.1 Background

The world has long committed to providing a quality education to every child. Beginning with the Universal Declaration of Human Rights in 1948, this commitment was reaffirmed in the Jomtien Declaration in 1990, the Millennium Development Goals in 2000 and again in the SDGs in 2015. During this period, school enrollment rates have risen dramatically in low- and middle-income countries, from less than half of children enrolled in school to nearly all children now attending at least some schooling.[22] Yet, learning has not kept pace and many children who are in school do not achieve even basic learning outcomes.[23]

Evidence has consistently shown that input-based approaches, such as providing textbooks or increasing teacher salaries, often do not yield expected results in improving educational outcomes.[24] Furthermore, interventions that are successful at small scales or when implemented by non-governmental organizations often do not achieve a similar impact when implemented at large scale within the government system.[25]

Despite inputs being a key part of an education system's production function, there is growing evidence pointing to the need for system transformation to achieve a quality education for every child.[26] This focus on system transformation is informed by a growing body of work on improving service delivery in public sector systems;[27] the importance of problem identification, iteration and adaptation for identifying effective policy approaches;[28] new approaches to bureaucratic accountability;[29] and the importance of system coherence and alignment around a common goal.[30]

---

[22] Barro, R. J., and Lee, J. W. "A new data set of educational attainment in the world, 1950–2010." *Journal of Development Economics* 104 (2013): 184–198.

[23] Pritchett, L. (2013). *The Rebirth of Education: Schooling Ain't Learning*. CGD Books.

[24] See for instance: Glewwe, P., Kremer, M., and Moulin, S. "Many children left behind? Textbooks and test scores in Kenya." *American Economic Journal: Applied Economics* 1 (2009): 112–135; de Ree, J., Muralidharan, K., Pradhan, M., and Rogers, H. "Double for nothing? Experimental evidence on an unconditional teacher salary increase in Indonesia." *The Quarterly Journal of Economics* 133 (2018): 993–1039; World Bank. *World Development Report 2018: Learning to Realize Education's Promise*. World Bank, 2018.

[25] Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., and Sandefur, J. "Experimental evidence on scaling up education reforms in Kenya." *Journal of Public Economics* 168 (2018): 1–20.

[26] Faul, M., and Savage, L. (eds). *Systems Thinking in International Education and Development: Unlocking Learning for All?* Cheltenham, UK: Edward Elgar Publishing; NORRAG Series on International Education and Development, 2023. https://www.elgaronline.com/edcollbook-oa/book/9781802205930/9781802205930.xml.

[27] World Bank. *World Development Report 2004: Making Services Work for Poor People*. Washington DC: World Bank, 2004. https://openknowledge.worldbank.org/entities/publication/c7755269-fa9d-5a07-9471-4ea35492b1d9.

[28] Andrews, M., Pritchett, L., and Woolcock, M. *Building State Capability: Evidence, Analysis, Action*. Oxford: Oxford University Press, 2017. https://bsc.cid.harvard.edu/publications/building-state-capability-evidence-analysis-action/; Watzlawick, P., Weakland, J. H., Fisch, R., and O'Hanlon, B. *Change: Principles of Problem Formation and Problem Resolution*. New York: W. W. Norton & Company, 2011. https://www.amazon.com/Change-Principles-Problem-Formation-Resolution-dp-0393707067/dp/0393707067/ref=dp_ob_title_bk; Crouch, L., and DeStefano, J. "Doing Reform Differently: Combining Rigor and Practicality in Implementation and Evaluation of System Reforms." RTI International Development Working Paper No. 2017-01. North Carolina: RTI International, 2017. https://www.rti.org/publication/doing-reform-differently/fulltext.pdf.

[29] Honig, D. "Managing for Motivation as Public Performance Improvement Strategy in Education & Far Beyond." CID Faculty Working Paper No. 409. Cambridge Massachusetts: Center for International Development, Harvard University, 2022. https://www.hks.harvard.edu/centers/cid/publications/faculty-working-papers/managing-motivation-public-performance.

[30] Pritchett, L. "Creating Education Systems Coherent for Learning Outcomes." RISE Working Paper Series.15/005. Oxford: RISE, 2015. https://doi.org/10.35489/BSG-RISE-WP_2015/005; Kaffenberger, M., and Spivack, M. "Chapter 7: System coherence for learning: applications of the RISE education systems framework." In *Systems Thinking in International Education and Development*, edited by M. Faul and L. Savage. Cheltenham, UK: Edward Elgar Publishing, 2023. https://www.elgaronline.com/edcollchap-oa/book/9781802205930/book-part-9781802205930-18.xml; Winthrop, R., and Sengeh, D. "Transforming education systems: Why, what, and how." Brookings Policy Brief.

The need for system transformation to improve educational outcomes has also been acknowledged and adopted by a growing group of global stakeholders. This was evidenced in the United Nation's Transforming Education Summit (2022);[31] joint statements by the World Bank, the United Nations Educational, Scientific and Cultural Organization (UNESCO), the United Nations Children's Fund (UNICEF), the Foreign, Commonwealth and Development Office (FCDO) and United States Agency for International Development (USAID);[32] and the Commitment to Action on Foundational Learning, endorsed by GPE among many others.[33] In 2020, GPE's Ministerial Advisory Group, made up of Education Ministers, called for GPE to "break from business as usual" approaches, and instead target transformative change.[34]

In this context, GPE's 2025 strategy embraces a system transformation paradigm as the means for ensuring a quality education for every child. GPE's operating model therefore aims to support countries to achieve system transformation in their context.

Figure 4, similar to figure 1 presented in section 2, presents the key features of the way GPE approaches system transformation, which involves empowering country-level actors to analyze systems, identify priorities, align actors and implement through iterative and adaptive approaches based on a shared diagnosis. This is meant as a radical departure from the approach of top-down planning, implementation and monitoring that has been common in sectoral coordination. This process forms the core of the GPE operating model (and therefore is the core object of this evaluation).
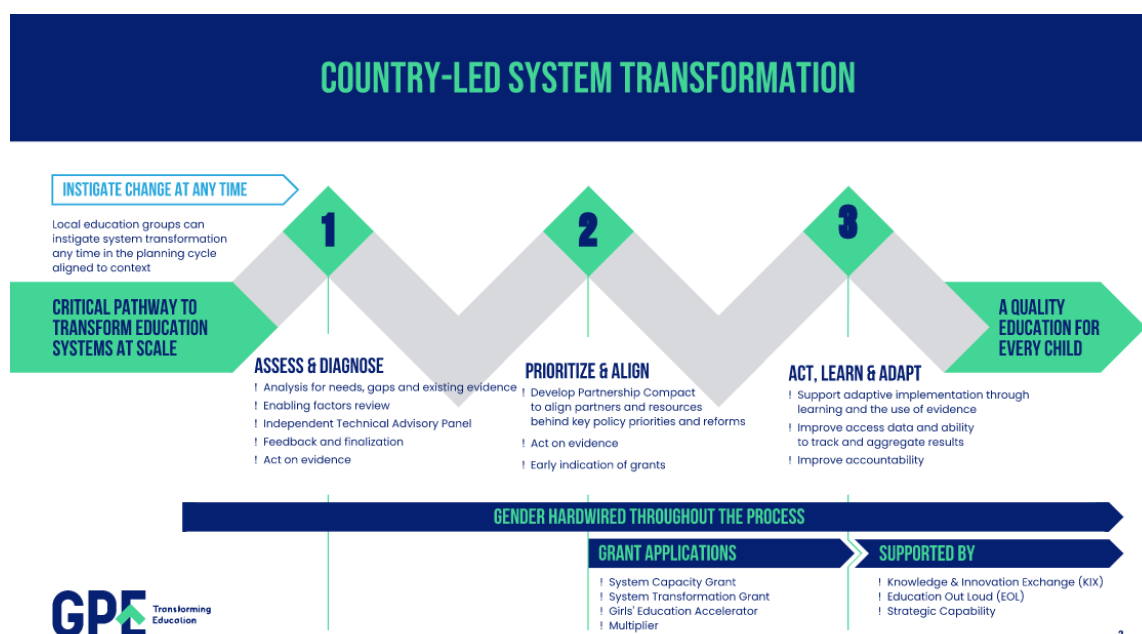
Figure 4. System transformation approach overview[35]

However, the adoption of new paradigms within an organization can be challenging. Operationalizing system transformation in the context of education reforms requires new conceptual models as in most

---

Washington DC: Brookings, 2022. https://www.brookings.edu/research/transforming-education-systems-why-what-and-how/.

[31] https://www.un.org/transforming-education-summit.

[32] see, for example, Saavedra, J. et al. "A new opportunity to help ensure all children learn and fulfill their potential." World Bank (blog), September 26, 2022. https://blogs.worldbank.org/education/new-opportunity-help-ensure-all-children-learn-and-fulfill-their-potential.

[33] https://www.unicef.org/learning-crisis/commitment-action-foundational-learning.

[34] https://www.globalpartnership.org/content/ministerial-aide-memoire-march-2020

[35] https://www.globalpartnership.org/what-we-do/transforming-education.

contexts it is a substantially different approach to change than business-as-usual approaches.[36] Preliminary consultations with the GPE Secretariat and other key experts on education system thinking suggest that there are varying degrees of understanding and alignment around the meaning of system transformation. Existing and varied interpretations of systems thinking might mean that GPE's proposed definition and approach may not always be universally embraced by other development partners, or may not be efficiently communicated, which can pose a challenge in the roll-out of the new operating model.

4.1.2 Key Concepts of System Transformation in the Context of GPE

The lack of agreement around defining and conceptualizing system transformation also poses a challenge for our evaluation, which at its heart seeks to understand the relevance, coherence, efficiency and effectiveness of GPE's support of countries' system transformation efforts. In this section, we put forward a set of definitions and concepts as interpreted by the evaluation team and which informs our understanding of GPE's 2025 strategy and operating model.

The definitions presented below are based on our review of the key literature and have been further refined to make them relevant to GPE's strategy and operating model. We also adapted the definitions for use in operationalizing our evaluation, for instance by ensuring coherence between key concepts. These definitions underpin the framework for our evaluation, which includes the theory of change for GPE's operating model, developed in consultation with the GPE Secretariat (and the assumptions within the theory of change) and the evaluation matrix.

We first offer a definition of system transformation itself, before defining additional concepts relevant to a system transformation paradigm. We define **system transformation** as a process through which education system stakeholders assess, diagnose and prioritize a key reform to catalyze learning improvements, align system actors and elements (including relevant subsystems) around the priority reform and implement the reform iteratively through learning and adaptation.[37] System transformation is a means to the end of improved learning and a quality education for every child.

- **Systems thinking:** involves interpreting a phenomenon as a system of dynamic, interconnected actors and elements. Rather than focusing on proximate causes of outcomes, systems thinking requires an understanding of the underlying drivers of observed outcomes. Applying systems thinking to education reforms requires diagnosing and addressing underlying drivers and constraints to the intended outcome (such as improved learning).[38]

- **System alignment**: is a key feature of the process of system transformation. This occurs when system actors, relationships and elements (including relevant subsystems) are in agreement on and oriented in the same direction to achieve a common goal, such as improving learning. Agreement on a goal is only the first step toward system alignment. Once a shared goal is established, actors, relationships, finance, information systems, support systems and other subsystems must also be oriented toward the same goal. In complex systems, the causal chains leading to goals are uncertain. Thus, in order to achieve alignment, experimental iteration is required. The GPE operating model seeks to promote systems alignment by creating processes and

---

[36] Examples include: (1) The Iceberg Model: Ecochallenge Dot Org. https://ecochallenge.org/iceberg-model/.; (2) The Cynefin Framework: Harvard Business Review, 2007. https://hbr.org/2007/11/a-leaders-framework-for-decision-making; (3) The Socio-Ecological Mode: Iacovidou, E., Hahladakis, J. N., and Purnell, P. "A systems thinking approach to understanding the challenges of achieving the circular economy." *Environmental Science and Pollution Research* 28 (2021): 24785–24806. https://doi.org/10.1007/s11356-020-11725-9. These are typically non-linear approaches to process diagnostics. Adopting them takes time and can require questioning more familiar linear models such as the logical framework.

[37] Adapted from GPE 2025 Strategic Plan and the RISE program: https://riseprogramme.org/publications/system-coherence-learning-applications-rise-education-systems-framework

[38] Adapted from GPE 2025 Strategic Plan and Oxfam: https://oxfamilibrary.openrepository.com/bitstream/handle/10546/579896/ml-systems-thinking-151020-en.pdf;jsessionid=D4DD54E01ECA407C1680778D60303E97?sequence=1)

incentives for system actors to agree on goals, develop reform plans based on these agreed goals and implement the reforms through a test–learn–adapt approach.[39]

- **Complexity:** in complex systems, the relationships and system elements are changed by their interactions with others in the system, in often unpredictable and unintended ways. This means that interventions cannot be planned reliably *a priori*. Instead planning must be continuously tested and involves assumptions that often begin as hypotheses. As education systems are (mostly) complex, linear theories of change do not capture the uncertainty of change. Highly iterative MEL feedback loops are needed to inform the implementation and adaptation of a reform on an ongoing basis.[40]

- **Enabling factors:** are system conditions that enhance system actors' capacity to achieve system transformation. Enabling factors take many forms, including system inputs, such as data used to inform the structure of a reform; outputs, such as data produced by effective monitoring systems; system processes, such as effective sector coordination; and other sector infrastructure.[41]

- **Experimental iteration:** with system transformation, and in order to achieve system alignment, the process of the implementation of a reform requires ongoing learning and adaptation. In complex systems, clear causal chains cannot be identified prior to implementation, which means ongoing experimentation and iteration is needed to achieve intended outcomes at scale.[42]

The concepts described above provide us with a defined vocabulary to evaluate the operating model's roll-out. To understand the effectiveness of GPE's efforts to support system transformation, the evaluation will seek to understand the extent to which country stakeholders undertook a systems-thinking approach to achieving transformation and the actual, or potential for, results and potential impact. We will explore whether there was an assessment and diagnosis of system drivers and constraints to change, whether actors successfully aligned around a common goal and prioritized reform, whether gender was hardwired and intersectional analysis was performed in the design phase of the reform, whether the reform is being rolled out in an iterative way to enable learning, adaptation and course correction, and the role played by GPE's support in the implementation of this process. Given the timing of the evaluation, we will not be able to evaluate outcomes and impacts of the reforms themselves (such as increased access to school or improved learning), but we will aim to assess likelihood that priority reforms are on track to produce intended outcomes based on reform-specific leading indicators of success.

4.1.3 Criteria for System Transformation in the Context of GPE and for this Evaluation

As indicated in section 3, evaluating system transformation is challenging. System transformation is a complex undertaking and it may take a long time for transformational outcomes to emerge. In addition to the use of different evaluation approaches (e.g., developmental evaluation), this evaluation also requires a consideration of different evaluation criteria to guide the evaluation.

The use of criteria is a critical feature of evaluation, which sets forth a standard or transparent norms for assessment. Evaluations are most prominently guided by the criteria set out by the Organisation for Economic Co-operation and Development's Development Assistance Committee (relevance, coherence, effectiveness, efficiency, impact and sustainability). However, as argued by Michael Patton,[43] traditional evaluation criteria alone are insufficient when examining transformative change, as they are focused on assessing the concept of a traditional "intervention" and therefore were not designed to capture the complexities of efforts required to support system transformation and the characteristics of transformed systems. Furthermore, scholars and evaluators studying and evaluating system

---

[39] Adapted from GPE 2025 Strategic Plan and the Brookings Institute:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3956220
[40] Adapted from the Brookings Institute: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3956220
[41] Adapted from GPE 2025 Strategic Plan and the Brookings Institute:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3956220.
[42] Adapted from the Education Commission (https://educationcommission.org/wp-content/uploads/2021/01/DeliverEd_Policy-Brief_1-1.pdf)
[43] See for instance, Patton, M. "Evaluation Criteria for Evaluating Transformation: Implications for the Coronavirus Pandemic and the Global Climate Emergency" *American Journal of Evaluation* 42 (2021): 53–89.
https://doi.org/10.1177/1098214020933689

transformation have recognized that evaluating system transformation requires the use of new or different criteria that are tailored to the context of system transformation.

Therefore, to support our evaluation, we developed a set of criteria for how we will assess system transformation in the context of GPE's operating model and strategy. These criteria are informed by the definitions and interpretations that we provided above, which will inform our conclusions as to whether the right principles of system transformation (in each context) are being applied in the design and implementation of the priority reform.

As each partner country can adapt its engagement with the GPE 2025 operating model in a way which is fit for its own purpose, it is also important that our criteria can also be adapted to help the evaluator understand what system transformation means in each context. Accordingly, the criteria set out below are guiding principles; in the spirit of our developmental evaluation approach, other criteria may emerge and be added during the evaluation.

As system transformation is a process, which includes the implementation of a priority reform, we have developed two sets of criteria: one related to the process that a country has undertaken to transform its education system, and a second related to the priority reform itself.[44] While it is beyond the scope of this evaluation to evaluate whether the intended reform outcomes have been achieved (e.g., improved learning outcomes), the criteria will help assess the likelihood that the reforms are on track to achieve their intended outcomes.

Criteria for the system transformation process include whether the following were carried out in quality, effective ways:

- A process of evidence-based **problem identification** was undertaken, to identify system-level constraints to the intended education outcome, and their root causes.

- A process of **evidence-review** was undertaken to determine evidence-informed approaches to addressing constraints.

- A process of **prioritization** was undertaken to establish the priority reform based on problem-identification and evidence-review processes, and accounting for resource and capacity constraints.

- A process of **stakeholder partnership and alignment** (including key education stakeholders) was undertaken to align around a priority reform and efforts to maintain stakeholder alignment are undertaken during reform implementation.

- A process of **iteration and adaptation** was built into the roll-out and implementation of the reform. This requires adequate data systems and processes for monitoring, evaluating and learning from outcomes (not just inputs) to understand effectiveness of reform and inform needed adaptations at regular intervals. By Phase 2, there is evidence that iteration and adaptation has occurred.

These criteria are mirrored in the operating model's compact process.

Criteria for system transformation reform include the following:

- The reform endeavors to achieve **improvements to learning with greater speed, scale and inclusion** than past and/or business-as-usual reforms.

- The reform endeavors to achieve learning improvements through approaches that are **evidence-based**.

- The reform **addresses multiple system constraints** through a multi-faceted approach to change.[45]

---

[44] The criteria have been developed building on the literature cited earlier in this section, with a particular emphasis on the following sources: GPE's description of transformation as depicted in https://www.globalpartnership.org/what-we-do/transforming-education); Andrews, M., Pritchett, L., and Woolcock, M. *Building State Capability*. UK: Oxford University Press, 2017; and Kaffenberger, M., and Spivack, M. "Chapter 7: System coherence for learning: applications of the RISE education systems framework" in *Systems Thinking in International Education and Development*, eds M. Faul and L. Savage. UK: Edward Elgar Publishing, 2023.

[45] For example, the Tusome Program in Kenya addressed multiple system constraints to learning by providing teacher training, ongoing teacher support through coaching, new teacher instructional materials, textbooks for every child and new mechanisms for monitoring and accountability, among others.

- The reform **aligns relevant subsystems, policies and practices**, such as teacher professional development, curriculum, assessment systems, education management and information systems (EMIS) and more, to achieve the intended outcomes.

- The reform aligns the incentives of actors from **all relevant levels and aspects of the education system** (e.g., national, regional and district).

- The reform includes mechanisms to ensure (and improve where needed) **implementation capacity and state capabilities**.

- The reform has been **adequately costed and sufficient financing** has been allocated.

- The reform includes **adequate and effective mechanisms to use data and evidence systems** to monitor, evaluate and learn from outcomes (not just inputs) and for iteration and adaptation (such as pre-defined intervals for reviewing outcomes data, identifying constraints to implementation and making adaptations accordingly).

These criteria provide a framework for assessing attributes of transformative reform. In our evaluation, we will examine how these criteria apply in each country's context. This is part of our investigation of the theory of change of the GPE model developed for our evaluation (see section 4.2). More specifically, this applies to how the criteria apply to events depicted under Phase 2 in the theory of change, as adapted to each country we investigate and has been embedded in our evaluation matrix.

4.2 Theory of Change

To support the evaluation's theory-based approach (outlined in section 3.2), we developed a theory of change to set out the way in which we interpret how the GPE operating model works to support country-level system transformation.[46]

The theory of change we developed differs from a more traditional program theory of change. It is particularly complex, with a strong focus on the GPE 2025 operating model itself (including the suitability of the model as a vehicle for facilitating change), as well as GPE's vision for country-level system transformation (Figure 4), which includes assessing and diagnosing system constraints; prioritizing and aligning around a selected reform; and acting on evidence, learning and adapting. As the GPE 2025 operating model aims to operationalize a system transformation approach, our theory of change provides a way of capturing the unfolding of this approach in practice.[47]

Our theory of change aims to map the plausible pathways toward the ultimate goal of the GPE operating model and strategy. In our theory of change, we have organized GPE's envisioned pathway to transformational change into four key processes and their associated outcomes: **(1) Developing a priority reform with transformation potential; (2) Operationalizing the priority reform; (3) Reaching results to the wider education system transformation through the reform;** and ultimately **(4) Achieving quality education for every child (SDG 4).**

Figure 5 presents the relationship between and progression of these four processes and also applies the "sphere of influence" framework, which helps set out the boundaries of control and influence that the GPE operating model and strategy have for each process.

---

[46] Our approach to developing our theory of change involved the following steps: input from the onboarding sessions in relation to the GPE 2025 model; the broader consultation sessions during the inception period; an examination of the existing compacts and draft compacts developed by partner countries and made available to the evaluation team; the workshop in Washington DC held on February 22–24, 2023 with GPE Secretariat staff and the external advisory panel members where a first draft of the theory of change and its assumptions was presented and discussed. Following the workshop, we further revised the theory of change with input from R&P to arrive at the version presented in this inception report.

[47] There are still ongoing adaptations to the GPE 2025 operating model that may affect the pathways within the theory of change for subsequent countries adopting the GPE model.
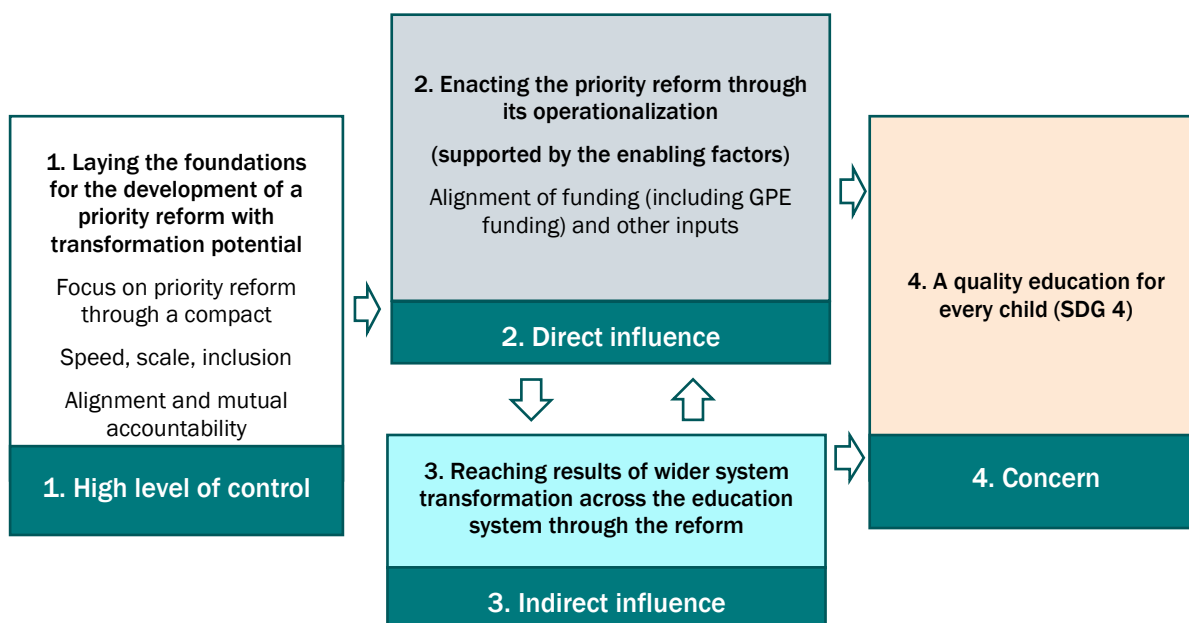
Figure 5. Depiction of the levels of control and influence that the GPE operating model and strategy have over the different levels of pathways toward transformational change

1. In the first process, the GPE 2025 operating model and strategy have a relatively **high level of control** and there is a higher degree of certainty (in this element) that the **development of a priority reform with transformational potential** will be achieved if the fidelity to the operating model principles and processes is maintained. This is a focus for the first phase of our country-level case studies, as the partner countries we are studying are only just beginning to implement their priority reforms.

2. The GPE operating model has a high level of **direct influence** on the **operationalization of the priority reform** through the provision of financial and non-financial support to country stakeholders. Changes at the country-level occur through iterative feedback loops (experimental iteration: act, learn and adapt).

3. The outcomes emerging as part of the priority reform are expected to lead to **wider education system transformation** through strengthened capacity and strengthened foundational conditions for sustaining the reform. The GPE operating model and strategy has only **indirectly influenced** the achievement of longer-term outcomes. There are feedback loops between the enactment of the priority reform and wider education system transformation.

4. Within the sphere of "concern," ultimately, the GPE operating model and strategy have a strong **commitment** to contributing to the achievement of **a quality education for all (SDG 4)** through the country-level efforts, and intends that support in steps 1, 2 and 3 above will contribute to this ultimate outcome. This indicates that while GPE is working toward the achievement of this ultimate outcome, it does not have control over this as a number of other external factors also affect this outcome.

These broad transformation expectations schematically summarized in figure 5 are shown in greater detail in our theory of change (Figure 6) and are explained below, which articulates our understanding of how the GPE operating model, including GPE's financial and non-financial support to reform broadly, aims to achieve impact.

Context - The conditions in which the program will operate (education system analysis, incl. political economy)

Goal: To accelerate (beyond incremental change) access, learning outcomes, and gender equality through equitable, inclusive, and resilient education systems fit for the 21st century

Phase 1 – Laying the foundations and preparing for the reform | **Phase 2 – Enacting the reform** | **Phase 3 – Reaching results through reform**

Inputs	Activities	Outputs	Outcomes	Impacts (Phase 1) - Inputs (Phases 2, 3)	Inputs	Activities	Outputs	Outcomes	Impacts (Phases 2, 3)
GPE functions fulfilled by **country-level actors** (MoE, CA, GA, LEG) and **Secretariat**	**Conduct pre-enabling factors review** e.g., convening, establishing working arrangements	Identified **critical bottlenecks** to system transformation	**Deeper understanding of the specific obstacles** to reform and agreement on solutions (learning from the system diagnostic process), including understanding how they will impact girls/boys differently	**Mobilized global/national partners and resources** for sustainable results (Enabling Outcome)	GPE functions fulfilled by **country-level actors** (MoE, CA, GA, LEG) and **Secretariat**	**Conduct relationship-altering initiatives** (dynamically connected actors driving and owning implementation and learning processes, under government leadership)	**Priority reform implementation produces desired outputs** at speed, scale, and with inclusion, mainstreaming gender equality	Improved quality, efficiency, and inclusiveness of education service delivery	Sustainable solutions at scale that accelerate a quality education for every child
GPE operating model tools for reform preparation, incl. guidelines, screening templates, applications, operating model videos, fact sheets, check lists, engagement memos, country analytics	**Diagnose and assess system constraints** to improvement though the enabling factors review (including gender-specific barriers to quality education faced by girls/boys)	Identified and agreed-upon **priority reform** in the partnership compact that is: (i) based on evidence and analysis of disaggregated data; (ii) addressing gender-specific barriers to quality education faced by girls and boys; (iii) costed; (iv) with aligned programs and agreed funding sources	**Improved policy making** through: (i) evidence-based problem identification, (ii) prioritization, (iii) agreement, and (iv) alignment to a reform which has the capacity to catalyze transformative change (elements of inclusion, speed, and scale), (iv) actions to address gaps in the enabling factors	**Strengthened gender responsive planning, policy development** for system wide impact (Country-Level Outcome 1)	**GPE operating model tools** for reform implementation, incl. guidelines, grant applications and templates, reporting templates, etc.			Improved equity and inclusiveness of learning outcomes	System-wide capacity strengthened across aspects of GPE country level objectives
Independent Technical Advisory Panel	**Review evidence** on effective (and cost effective) policies and reforms to address identified constraints			**Mobilized coordinated action and financing** to enable transformative change (Country-Level Outcome 2)	**GPE country-level financial support** to reform implementation (e.g., STG, MLT, GEA, PDG) and technical assistance (e.g., strategic capabilities)	**Establish systems for ongoing disaggregated data collection and gender analysis** of outcomes		Reduced gender-specific barriers to a quality education for girls/boys	
Secretariat quality assurance function and related support for reform preparation	**Negotiate and conduct policy discussions**		**Improved gender transformative approaches** to planning, monitoring, and policy development	**Strengthened capacity** to adapt and learn, to implement, and drive results at scale (Country-Level Outcome 3)	**Secretariat quality assurance function** and related support for reform implementation	**Act, adapt from continuous learning** from implementation: (i) new/evolving understanding of obstacles to reform; (ii) Identification and scaling of successful practices		Improved access to a quality education for all girls/boys	
Development partners' existing policy processes, funding, support for reform preparation	**Build coalition** between stakeholders around a chosen policy, including to improve gender equality in/through education			*These impacts will continue to unfold post compact and feed into conditions for success for Phases 2 and 3*	**Education Out Loud**		**Gaps in the enabling factors are addressed** through capacity and policy-related reforms	**Strengthened organizational capacity** through reform / Improved gender-aware policy-making, management, and governance of the education system	
	Align commitments and resources (incl. programs, funding, technical assistance) among stakeholders around a chosen policy	Inclusive (inter)sectoral **dialogue** on bottlenecks, reform, and alignment	**Efficient coalescing of programming and funding** which are aligned to the priority reform		**Development partners'** existing policy processes, funding, support				
Global, regional inputs to reform preparation: Evidence and related networks (including KIX), advocacy, technical assistance, funding	**Connect stakeholder groups**, i.e., understand and mediate dynamics between change actors; empower lesser-heard voices at country level		**All partners** (within and outside sector) **work better together under government leadership and ownership**, working together and holding each other mutually accountable		**Global, regional inputs** to reform implementation: Evidence and related networks (including KIX), advocacy, technical assistance, funding	**Address differentiated impacts on girls/boys** from reform implementation		**Strengthened foundational conditions for reform** (domestic finance; data and evidence; gender-responsive planning and monitoring; sector coordination)	

System Capacity Grant (SCG) can be accessed at any time (alongside funding from other sources)
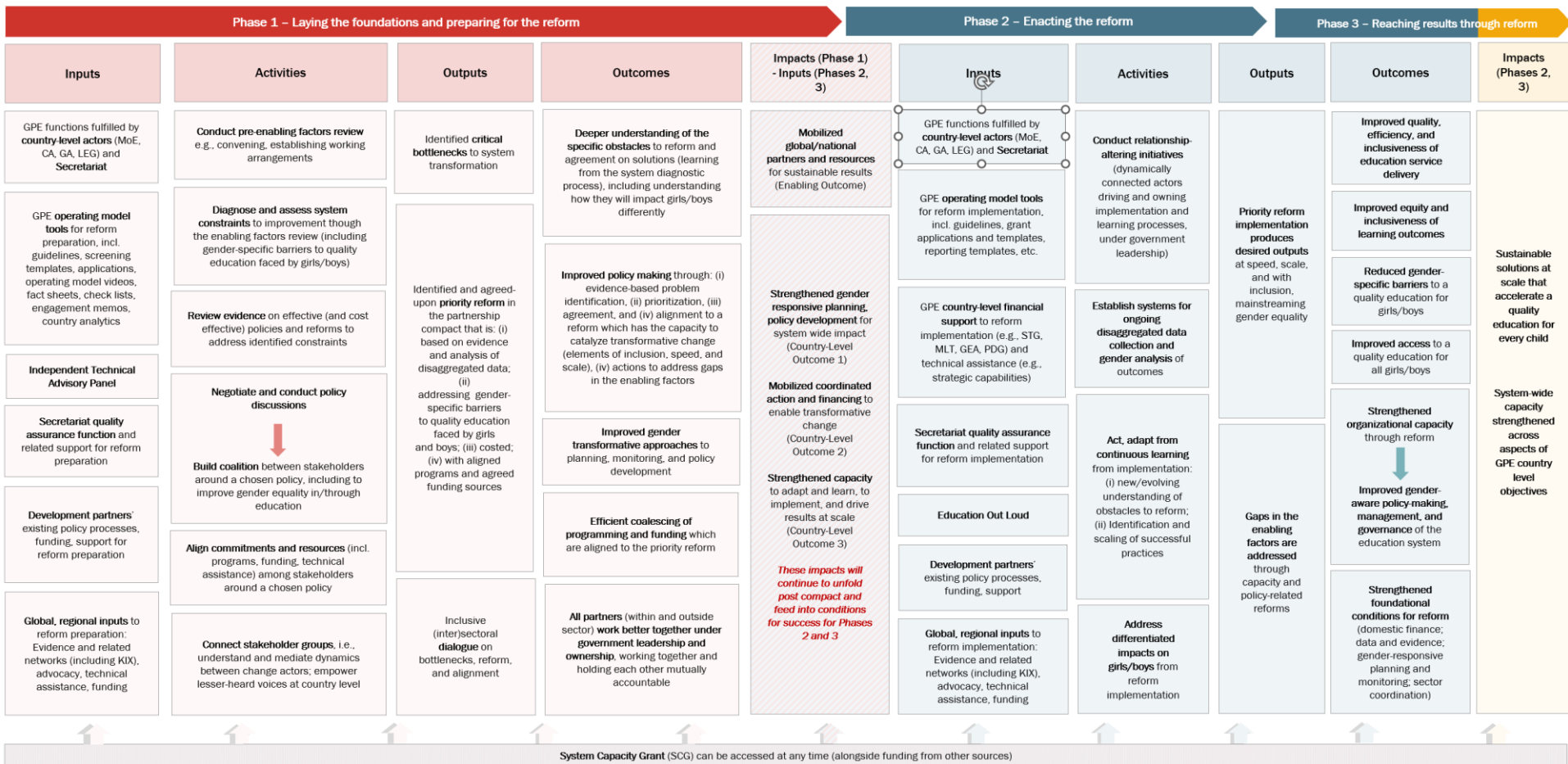
Figure 6. Theory of change for the Country-Level and Thematic Evaluation

Straddling our theory of change is the **context**. The theory of change acknowledges that the direct influence of GPE's operating model and strategy diffuses as we move toward impact. There will be many contextual factors and also assumptions (see below) which will influence its success. Therefore, throughout the theory of change and importantly as events move toward impact, it will be important to capture the conditions within the broad social, political, economic and environmental landscape which either support or hinder the reform efforts. This includes the political economy, national development visions and urgent problems facing countries, for example. This is a critical area for our evaluation to explore, utilizing our realist evaluation approach.

The **goal** referenced in the theory of change is the goal set out in the GPE 2025 Strategic Plan.

Phase 1 in the theory of change, **"Laying the foundations for priority reform,"** outlines the expectations of the process of identifying and designing a priority reform and compact development for a country. This leads to the expected impacts as set out in the GPE 2025 Strategic Plan and Results Framework: Mobilized global/national partners and resources for sustainable results; Strengthened gender-responsive planning, policy development for system-wide impact; Mobilized coordinated action and financing to enable change; and Strengthened capacity to adapt and learn, to implement and drive results at scale. These impacts are expected to continue to unfold in Phases 2 and 3.[48]

Phase 2 in the theory of change, **"Enacting the Reform,"** represents the activities and outputs of the priority reform itself, recognizing the system-level levers of relationship-altering initiatives, systems to collect data, acting and adapting for continuous learning and addressing differentiated impacts on girls and boys. This leads to change that respects the principles of speed, scale and inclusion and the continued need to address enabling factors. Phase 2 will differ in each country, based on the choice of priority reform and the way in which system transformation is understood by the stakeholders.

Phase 3 in the theory of change, **"Reaching results through reform,"** links the results (outputs and outcomes) of the priority reform to wider system transformation leading to sustainable solutions at scale that accelerate a quality education for every child and system-wide capacity strengthening across aspects of the GPE country-level objectives.

The theory of change also accounts for GPE's strategic aim of promoting **gender equality to, within and through education**. In Phase 1, the assessment of the enabling factors is key to determine gender-based challenges and constraints to access to quality education for girls and boys. This allows for the alignment of stakeholders and wider coalition around selected reforms that tackle (directly or indirectly) gender barriers to quality education. The expected outcomes and impacts of strengthened gender-responsive planning are the foundations of Phases 2 and 3, where collecting and analyzing disaggregated data will allow for adaptation of activities during the implementation of the reform and continuous learning. This is necessary to address gaps in access to quality education and ultimately reach the goal of gender-transformative policies and practices in education systems.

Our evaluation has greater control over assessing the first two phases, where GPE has most control or direct influence, where we are able to assess both the process of and outcomes resulting from the roll-out of the operating model and the operationalization of the priority reform. This forms the focus of the evaluation throughout the baseline and midline phases of our country-level case studies, while in later phases (midline and endline) we will begin to investigate the achievement of or potential for outcomes in the latter two processes.

It is important to note that this version of the theory of change is a broad tool that allows us to understand and interpret changes at the country-level. However, it is unlikely to apply in its entirety in any given country. For each of the country-level case studies, we will use this version of the theory of change as a starting point to develop a country-specific theory of change, in which we will map their expected pathway to impact on this overarching theory of change, and articulate the expected inputs, outputs and outcomes that are specific to each country.

---

[48] The evaluation also incorporates a thematic lens with a focus on gender and domestic financing. For more information on the thematic evaluation aspect and the phases see section 3.2.

4.2.1 Assumptions Embedded in the Theory of Change

Assumptions are assertions made about key aspects of a theory of change that underlie the plan and which, if they do not hold true, negatively affect the change pathway and therefore the delivery of the anticipated activities, outputs, outcomes and impacts.[49]

Below, we set out the assumptions that are embedded within, but not necessarily articulated in, our theory of change. These assumptions are not explicitly addressed in the program design but nevertheless are expected to influence the delivery of the priority reform. These assumptions are important to test in a theory-based and realist evaluation as they represent the potential contexts, circumstances, underlying conditions or resources that need to exist for planned change to happen.

There is a set of seven assumptions that cut across the theory of change and remain valid for all parts of the program. We have articulated an additional four assumptions which are specific to particular phases of the theory of change. Finally, the assumption that gender equality to, within and through education is a desirable change not only for GPE but also for all stakeholders involved in the system transformation process, is transversal to all assumptions.

Assumptions that cut across the theory of change

1. The composition of the **local education group** is inclusive of different voices, including of civil society and teacher groups, and women's interests representatives within.

2. A critical mass of **country-level partners** (including Grant Agents and Coordinating Agencies) is active and remains fully **committed** to and **demonstrates** (a) working together, under government leadership, during the compact development process and operationalization of the priority reform (including by avoiding aid fragmentation), (b) learning together/continuously adapting and (c) being held accountable (including around gender mainstreaming).

3. At country level, there are appropriate incentives, **functioning mechanisms and sufficient capacity** (or plans to address capacity gaps) (including GPE actors such as Grant Agents and Coordinating Agencies) at the country level to (a) co-develop, operationalize the priority reform, (b) avoid fragmentation of aid, (c) learn together/continuously adapt and (d) be held accountable (including around gender mainstreaming). Additional required support is available in **partner countries affected by fragility and conflict**.

4. The **GPE Secretariat** remains **committed** to the GPE 2025 operating model and gender hardwiring within the model, and to continuously learning/adapting as a means of supporting wider education system transformation, and it has the **capacity** to support its roll-out and implementation.

5. After accounting for GPE grants and funding from other donors, **domestic finance** remains adequate to deliver priority reforms, including dedicated budget for gender reforms.

6. There are signs that beliefs and perceptions concerning the most important **purpose of school** (and education in general, especially around gender equality in education) are converging among the relevant stakeholders and align to the objectives of the priority reform.

7. There are functioning mechanisms in place to address usual **resistance to reform,** including resistance to transformation that challenges gender roles, norms and stereotypes within society.

Assumptions that apply to Phase 1 – Laying the foundations and preparing for the reform

8. The **enabling factors, ITAP processes, compact development processes** (including data disaggregation and gender-responsive sector planning, policy and monitoring) and **SCG** are well **understood**, considered **relevant** and **sufficient** as a means of solving bottlenecks and strengthening capacity; and **enable good policy dialogue and capacity building** on system capacity gaps.

Assumptions that apply to Phase 2 – Enacting the reform, and Phase 3 – Reaching results

9. The **GPE support and related mechanisms** mobilized to facilitate reform implementation and continue capacity strengthening (e.g., STG, GEA, MLT and SCG) and other compact partners' support, are well

---

[49] Assumptions form an important part of theory-based and developmental evaluations.

understood and considered **relevant, sufficient and cost-effective** as a means of implementing the priority reform.

10. **Adequate and timely data/evidence** at the country-level is **available** and appropriately **used** to allow for monitoring and adaptation during the operationalization of the compact and priority reform, including the implementation of the enabling factors and the priority reform (including gender-disaggregated data, gender analysis and intersectional factors analysis).

11. GPE support is being adapted based on **emerging emergency circumstances** (such as natural disasters, health emergencies and conflict).

We will explore these assumptions during interviews as part of our country-level case studies. A full overview of the assumptions, indicators, data sources and analytical tools is found in **Annex 2**.

4.3 Evaluation Matrix

An evaluation matrix serves as a guide for the evaluation set-up and operationalization. For this evaluation, we developed our evaluation matrix in close collaboration with the R&P team.

In our evaluation matrix, the evaluation objectives are articulated as **key leading evaluation questions**, which frame the inquiry of the evaluation. The evaluation questions were designed to capture how we will investigate and interrogate the theory of change set out above. The evaluation questions also go beyond this, and capture other learning needs, as framed by the objectives of the evaluation. The evaluation questions are then broken down into **subquestions**, which provide further interpretation of the key leading evaluation question and help with its operationalization. We also make use of **prospective** evaluation questions and subquestions, which allow us to look ahead and anticipate the likelihood of change, in instances in which the evaluation time frames mean we are unable to allow enough time for outcomes to be realized.

We have also included **indicators**, which explain how we expect to use the evidence collected to reach judgments against the subquestions. Finally, in our evaluation matrix we describe **data sources and methods** indicating what data we will collect and how we will analyze them. In order to reduce bias as much as possible, each evaluation question is expected to be answered using evidence from multiple data sources which will be triangulated to ensure that the evaluative judgments are robust.

Presently, our evaluation matrix reflects the focus of the evaluation for our baseline studies (of our country-level case studies and thematic case studies on domestic financing and gender equality). These will apply in Phase 1 for our eight Wave 1 case studies and in Phase 2 for our seven Wave 2 case studies. We will further develop our evaluation matrix to guide the midline and endline studies, which will be developed at the start of Phases 2 and 3, respectively (and outlined in our phase concept notes).

Below, we present the evaluation matrix for the evaluation.

Leading question	Subquestion	Indicators	Data sources and analysis
		Baseline (country-level case studies)	
1. Do the reforms that partner countries have prioritized in their partnership compacts, and their associated pathways of change, demonstrate potential to transform their education systems? What makes these reforms transformative?	Construct validity 1.1 To what extent, and how, is system transformation being understood and conceptualized in different country contexts? 1.2 Do the reforms that partner countries have prioritized in their partnership compacts meet the criteria of system transformation process and of transformation reform? 1.3 To what extent is gender equality hardwired in these reforms?	1.1 Key country-level stakeholders (a) understand/agree with the need for system transformation (e.g., need for improved learning and equity at scale, accelerated progress in outcomes, including the most marginalized etc.) (qualitative data); (b) understand and align on what it concretely means in their context (not a theoretical definition but, for example, what the obstacles to reform are, what kinds of reforms can be transformative etc. in their own country); (c) understand/agree with the need for gender hardwiring and what it concretely means in their context; and (d) understand and are committed to the systems reform process. 1.2 Prioritized reforms in partnership compacts meet the criteria of systems transformation process and of transformative reforms (e.g., addressing the root causes of system underperformance/obstacles to reform; prioritized; potential for impact at scale and/or with speed and inclusion; and alignment of necessary stakeholders). 1.3 Gender equality considerations related to the prioritized reforms were analyzed, based on GPE's approach, and appropriate actions to address them have been embedded in the priority reforms.	1.1 Identification of a set of attributes of transformative reform based on a review of the literature on systems change in education and GPE documentation on the new strategy and operating model, including approach to/definition of transformation. 1.2 Key-informant interviews and/or focus group discussions with country-level stakeholders and Secretariat staff and qualitative content analysis. 1.3 Desk review of prioritized reforms in partnership compacts against attributes of transformative reform (with an appropriate review of supporting documentation and data like the enabling factors analyses, education sector analyses, EMIS data and education sector plans to contextualize this review) and against attributes of gender equality also defined based on GPE's approach and the literature on gender equality in and through education.
	Operational validity 1.4 How do reforms that partner countries have prioritized in their partnership compact compare to previous reforms in that area (including with respect to the attributes of transformative reform)? How well do	1.4 Prioritized reform considers evidence on past constraints to reform, and includes policies and programs to address those constraints. 1.5 Compared to previous reforms related to the prioritized reforms (e.g., reforms that aimed to achieve the same/similar objectives), the prioritized reforms more strongly demonstrate the attributes of transformative reforms (or improve the design/implementation arrangements/degree of	1.4 Comparative desk review of prioritized reforms in partnership compacts against previous related reforms, based on the attributes of transformative reform, system-level change and gender equality hardwiring criteria (with an appropriate review of supporting documentation and data like the enabling factors analyses, education sector

Leading question	Subquestion	Indicators	Data sources and analysis
	the reforms fit within the countries' broader education plans and policies?	stakeholder alignment of previous transformative reforms).	analyses, EMIS data and education sector plans to contextualize this review).
	1.5 Are the theories of change underlying the prioritized reforms (whether explicitly defined or not), and related assumptions, credible and suitably informed by evidence about the bottlenecks, the need for the reform and policies and interventions to address them?

1.6 Do the theories of change (whether explicitly defined or not) identify pathways to address the gender-related barriers to education faced by girls and boys associated with the reforms? Are these pathways credible and suitably informed by evidence?

1.7 To what extent are the enabling factors for system transformation (gender-responsive planning and monitoring, data and evidence, sector coordination, and domestic finance) present? Where there are gaps in the enabling factors, are the proposed policy actions to address these gaps adequate to enable transformative reform? Are the actions credible and suitably informed by evidence? | 1.6 Theories of change (and associated assumptions) of the prioritized reforms, as implicitly or explicitly described in partnership compacts, are valid; interventions to address gender-related barriers to education faced by girls and boys are evidence-based and embedded in the theory of change.

1.7 The extent to which the identification/development of the prioritized reform considered contextual factors that will affect reform credibility (such as timing of policy cycle, financing gap, previous lessons learned, system governance and stakeholder incentives).

1.8 Degree of alignment between the gaps identified in the enabling factors and the proposed policy actions to address them; suitable evidence is cited to demonstrate that the proposed policy actions will address the gaps identified. | 1.5 Desk-based assessment of partnership compacts to assess (a) the degree of alignment between prioritized reforms in partnership compacts and sector plans and policies; (b) extent to which theories of change are evidence-informed and include suitable pathways to address gender-related barriers; (c) extent to which policy actions to address gaps in the enabling factors are aligned with the gaps identified through the enabling factors analysis and are suitably informed by evidence; and (d) gender hardwiring.

1.6 Key-informant interviews and/or focus group discussions with country-level stakeholders, ITAP and Secretariat staff and qualitative content analysis. |
| | Alignment

1.8 To what extent are key education stakeholders aligned with the reforms prioritized by partner countries and the policy actions to address the enabling | 1.9 Degree of alignment among key education stakeholders (defined as those who are necessary for the success of the reform) and their programs/resources behind the prioritized reforms and actions to address the enabling factors. Degree of alignment between | 1.7 Desk review of partnership compacts to identify partner commitments and resources/programs mobilized to align with the prioritized reforms and/or policy actions on the enabling factors. |

Leading question	Subquestion	Indicators	Data sources and analysis
	factors for transformative reform? How effective was the partnering process that led to these reforms and actions being chosen? Have partnerships/partnering processes at the country level changed over time (especially partnership around gender equality)?	institutional mechanisms and country goals, alignment of education policies and practices. 1.10 Extent to which key education stakeholders agree that the dialogue leading to the prioritized reforms and policy actions was national government-led, inclusive of key education actors, evidence-based and promoted mutual accountability for the implementation and achievement of the prioritized reforms.	1.8 Survey of country-level stakeholders and quantitative analysis of the survey results. 1.9 Key-informant interviews and/or focus group discussions with country-level stakeholders and Secretariat staff and qualitative content analysis. 1.10 Political economy analysis.
	Learning 1.9 Are the pathways in the theory of change identified clearly enough to guide implementation of the prioritized reforms and to monitor, evaluate and learn from implementation? Are there adequate provisions (MEL frameworks; mechanisms for joint review and learning by partners etc.) and suitable conditions (e.g., policies, capacities and partner interest) for joint monitoring, evaluating and learning from the implementation of the prioritized reforms? Have the countries' approaches to data and learning changed over time (especially data/learning around gender equality)?	1.11 Objectives of the pathways in the theory of change are "SMART" (e.g., specific, measurable, achievable, relevant and time-bound). 1.12 Adequate MEL frameworks and plans have been developed (or there are plans for these to be developed) for the prioritized reforms (whether described in the partnership compact or in other documents that describe the policies and programs underlying the prioritized reforms). 1.13 Adequate mechanisms and conditions for joint monitoring and evaluation of the implementation of the prioritized reform are in place/being put in place (with the participation of the necessary stakeholders). 1.14 Suitable mechanisms and conditions for learning from implementation and adaptation are in place/being put in place (e.g., timely evidence, sustained government and partner interest and capacities for learning).	1.11 Desk review of the partnership compacts (and documents related to the policies and programs underlying the prioritized reform in the partnership compact). 1.12 Review of documentation related to past and planned joint monitoring and evaluation exercises/mechanisms in place for policies and programs. 1.13 Key-informant interviews and/or focus group discussions with country-level stakeholders and Secretariat staff and qualitative content analysis.
2. How relevant, efficient, effective and coherent is GPE support in helping countries to identify transformative reforms and align	Facilitating system transformation 2.1 How well did GPE support countries to conduct policy reform and implement	2.1 Extent to which enabling factors self-assessments identify the key bottlenecks to system transformation in the areas covered by the enabling factors; are evidence-based, of quality, and helpful for informing policy dialogue; are feasible given country capacity; and country-level stakeholders believe that the value added	2.1 Criteria-based desk reviews of enabling factors analyses, ITAP assessments, partnership compacts and SCG program documents; review of Board decision documents (with an appropriate review of supporting documentation and data such as the enabling factors analyses, education sector

Leading question	Subquestion	Indicators	Data sources and analysis
partners and resources to them?	sector programs under GPE 2020[50] (especially the support received from the grants)?

2.2 As operationalized, is the GPE operating model efficiently and effectively helping countries, in different contexts, to (a) identify bottlenecks to system transformation, (b) ensure that the enabling factors (and related capacity) for system transformation are in place, (c) identify and prioritize reforms that have potential for system transformation and (d) align partners and resources around countries' prioritized reforms?

2.3 Is GPE support (a) aligned with the countries' own processes and policy cycle and (b) tailored and adaptive to country contexts and needs (e.g., countries where existing reform priorities and their objectives are well defined, countries where a broad set of plans and policies exist but priorities are not well defined, and countries where plans and policy frameworks are weak)? How have country context and explicit (including adaptations to model processes made by the Secretariat)/implicit variations in GPE support affected the relevance, | by the enabling factors self-assessments justifies the cost of preparing them.

2.2 Extent to which ITAP assessments are robust and evidence-based; fulfill their role of providing independent, technical assessments of country status with respect to the enabling factors areas; are consistent across countries; and inform country-level policy dialogue.

2.3 The top-up portion of the STG served as an incentive, to address gaps in the enabling factors.

2.4 The enabling factors self-assessments and compact processes strengthening the capacity of country-level stakeholders to diagnose system bottlenecks, identify transformative reforms and align stakeholders behind these reforms.

2.5 The SCG is providing capacity support that is aligned with country needs and likely to strengthen country capacity.

2.6 The partnership compact processes supporting/leading to greater stakeholder alignment behind prioritized reforms.

2.7 Extent to which partnership compact processes were (a) aligned with own country cycles and processes, (b) adaptive to context; feasible given country capacity and (c) stakeholders believe the value added by compacts justify the cost of preparing them.

2.8 GPE operating model processes and support (including enabling factors self-assessments, compact processes, top-up portion of the STG, GEA and MLT) incentivized and/or supported countries to consider and | analyses, EMIS data and education sector plans to contextualize this review).

2.2 Assessment of the alignment of documented operating model adaptations made by the Secretariat to challenges identified by stakeholders and their root causes.

2.3 Survey of country-level stakeholders and quantitative analysis of the survey results.

2.4 Key-informant interviews and/or focus group discussions with country-level stakeholders, ITAP, staff of partners in regional or global offices and Secretariat staff, and qualitative content analysis.

2.5 Comparisons to GPE 2020 support will be based primarily on document reviews, particularly grant documents. |

---

[50] GPE Strategic Plan 2016-2020: GPE 2020 Improving learning and equity through stronger education systems.
https://www.globalpartnership.org/node/document/download?file=document/file/2018-05-gpe-2020-strategic-plan.pdf.

Leading question	Subquestion	Indicators	Data sources and analysis
	efficiency, effectiveness and coherence of GPE's support?	address gender equality in the prioritized reforms and policy actions on the enabling factors.	
	2.4 Does the operating model effectively support and incentivize partner countries to hardwire gender equality into their prioritized reforms?	2.9 Adjustments to the GPE model made since its launch were based on suitable evidence, were appropriate based on the causes of the shortcomings identified and ultimately remediated these causes; extent to which the adaptations to the operating model implemented during the operating model roll-out correspond to the challenges experienced by stakeholders (as identified by the stakeholders).	
	2.5 Has the GPE model evolved/been adapted since GPE 2025 was launched, and if so, on what basis and how successfully?		
	2.6 What are positive or negative unintended consequences of the operating model?		
	Supporting the partnership at country level	2.10 See 1.10 above.	2.6 Survey of country-level stakeholders and quantitative analysis of the survey results.
	2.7 Is the operating model supporting inclusive policy dialogue, government leadership, and mutual accountability toward commitments made by partners in support of the priority reform? Are the different GPE actors fulfilling their expected roles to support the country (that is, Secretariat, Coordinating Agency and Grant Agent)? How has this evolved over time (especially around gender equality)?	2.11 Extent to which stakeholders believe GPE actors are fulfilling their expected roles to support the country.	2.7 Key-informant interviews and/or focus group discussions with country-level stakeholders, ITAP, staff of partners in regional or global offices and Secretariat staff, and qualitative content analysis.
		2.12 Level of country-level stakeholder satisfaction with Secretariat support; clarity and consistency of operating model guidelines and templates.	
		2.13 GPE operating model processes and support strengthened the partnership at country level with respect to inclusive policy dialogue, government leadership, partner alignment and mutual accountability (especially around gender equality) (e.g., relevant civil society organizations and other groups consulted, and greater depth of participation).	2.8 Desk review of partner commitments in partnership compacts. 2.9 Political economy analysis.
	Supporting learning and adaptation	2.14 Extent to which GPE processes and support help establish the provisions and mechanisms for joint MEL by partners (especially around gender equality).	2.10 Survey of country-level stakeholders and quantitative analysis of the survey results.
	2.8 To what extent is the operating model supporting countries to put adequate provisions (e.g., MEL frameworks; and mechanisms for joint		2.11 Key-informant interviews and/or focus group discussions with country-level

Leading question	Subquestion	Indicators	Data sources and analysis
	review and learning by partners) and suitable conditions (e.g., policies, capacities and partner interest) for joint MEL from the implementation of the prioritized reforms in place (especially around gender equality)? Has this evolved over time?		stakeholders and Secretariat staff and qualitative content analysis. 2.12 Desk review of partnership compacts, SCG program documents.
3. Are partner countries well-positioned to implement their transformative reforms? How is GPE supporting/ contributing to position countries to implement their transformative reforms well?	3.1 Are the necessary capacities, resources and mechanisms (a) identified and (b) in place or planned for, to ensure the prioritized reforms will be implemented? 3.2 Which factors support the implementation of the transformative reform in the country? Which risks may threaten implementation, and are they being addressed adequately? Does GPE support ensure the supportive factors are leveraged, and risks anticipated and dealt with to implement the chosen reforms? 3.3 Given their characteristics, what is the potential for success of the partnership compacts in achieving the prioritized reforms? What should be modified now to improve the future chances of success of the priority reforms? 3.4 What other support does GPE provide for long-term continuity of the reform?	3.1 Extent to which prioritized reforms are costed and financed, with appropriate implementation plans in place/or there is evidence that such costing, financing and implementation planning will be carried out (either in the partnership compact or in documents related to the programs and policies that underlie the prioritized reform). 3.2 Extent to which stakeholder roles and responsibilities in implementing the prioritized reforms are clear; the respective stakeholders are aligned with their expected roles and responsibilities, and have the capacity to fulfill them. 3.3 Appropriate mitigation actions were taken to address risks to the implementation of the prioritized reforms and policy actions on the enabling factors. 3.4 Extent to which GPE processes and support are adequate and sufficient for the implementation (in addition to domestic resources and capacity) of the prioritized reform and/or help countries to identify and leverage additional resources where necessary.	3.1 Survey of country-level stakeholders and quantitative analysis of the survey results. 3.2 Key-informant interviews and/or focus group discussions with country-level stakeholders and Secretariat staff and qualitative content analysis. 3.3 Desk review of partnership compacts.
		Baseline (domestic finance thematic case study)	

Leading question	Subquestion	Indicators	Data sources and analysis
4. What is the status of partner countries with respect to domestic finance?	4.1 What is the status of partner countries with respect to domestic finance? 4.2 What is the level of government and partner capacity to identify and address system bottlenecks with respect to domestic finance?	4.1 GPE results framework indicators on domestic finance; enabling factors "guiding considerations" on domestic finance; other standard international indicators. 4.2 Extent to which country-level stakeholders have/report having sufficient capacity and access to data to carry out diagnostics on domestic finance (including the enabling factors self-assessment) and develop solutions to address the gaps identified.	4.1 Desk review of domestic finance diagnostics (including enabling factors self-assessments, ITAP assessments and country responses to ITAP assessments) and domestic finance-related policy actions in partnership compacts; collection and analysis of domestic finance-related data from other standard sources (UNESCO Institute for Statistics – UIS; education sector analyses etc.). 4.2 Analysis of domestic finance-related data gathered through the country-level case studies' data collection. 4.3 Interviews with Secretariat domestic finance specialists and qualitative content analysis.
5. What, if any, policy actions are partner countries choosing to undertake with respect to domestic finance? Are these policy actions likely to result in the necessary domestic finance-related enabling conditions for transformative reform to be in place?	5.1 What domestic finance-related policy actions are partner countries undertaking? 5.2 Are the chosen policy actions aligned with the domestic finance challenges the countries face? 5.3 To what extent does the successful implementation of the policy actions have potential to address the challenges?	5.1 Extent to which the policy actions proposed in the partnership compacts are evidence-based (including being aligned with gaps identified through the enabling factors self-assessments), appropriately designed and sufficiently ambitious to address the challenges identified.	5.1 Desk review of domestic finance diagnostics (including enabling factor self-assessments, ITAP assessments and country responses to ITAP assessments) and domestic finance-related policy actions in partnership compacts; collection and analysis of domestic finance-related data and policy documents from other standard sources (UIS, education sector analyses etc., government policy documents). 5.2 Analysis of domestic finance-related data gathered through the country-level case studies' data collection. 5.3 Interviews with Secretariat domestic finance specialists and qualitative content analysis. 5.4 Interviews with ministries of finance

Leading question	Subquestion	Indicators	Data sources and analysis
6. What support, with respect to domestic finance, is GPE providing to partner countries to design and align resources to their chosen policy actions? How relevant, efficient, effective and coherent is GPE's financial and non-financial support in helping countries to identify and address system bottlenecks with respect to domestic finance?	6.1 To what extent did the enabling factors self-assessments and compact dialogue (including country self-assessment and ITAP assessment) support the effective identification of system bottlenecks in domestic finance? What aspects/features (e.g., process steps, methodology and guidance) of the process were most effective? How efficient was the process? 6.2 To what extent has GPE's allocation mechanism (minimum allocation + top-up) served as an incentive to address system bottlenecks in domestic finance? 6.3 To what extent did the partnership compact process align partners behind the chosen policy actions with respect to domestic finance? 6.4 To what extent did the SCG support partner country capacity to address domestic finance-related system bottlenecks? Was the capacity support provided aligned with country needs?	6.1 Extent to which enabling factors self-assessments identify the key bottlenecks in domestic finance are evidence-based, of quality and inform policy dialogue; feasible given country capacity; country-level stakeholders believe that the value added by the domestic finance self-assessments justifies the cost of preparing them. 6.2 Extent to which ITAP assessments of domestic finance are robust and evidence-based; fulfill their role of providing independent, technical assessments of country status with respect to the enabling factors areas; consistent across countries; inform country-level policy dialogue. 6.3 Extent to which the top-up portion of the STG served as an incentive to address gaps in domestic finance. 6.4 Extent to which the partnership compact processes supported/led to greater stakeholder alignment behind policy actions to address gaps in domestic finance. 6.5 The enabling factors self-assessment and compact processes strengthened the capacity of country-level stakeholders to diagnose domestic finance-related system bottlenecks, identify corresponding policy actions and align stakeholders behind the policy actions identified. 6.6 Extent to which domestic finance-related activities financed through the SCG are aligned to country needs and are likely to strengthen country capacity.	6.1 Desk review of domestic finance enabling factors assessment documents based on criteria corresponding to the indicators/success criteria (including ITAP assessments and country responses to ITAP assessments) and domestic finance-related policy actions in partnership compacts; collection and analysis of domestic finance-related data and policy documents from other standard sources (UIS, education sector analyses etc., government policy documents). 6.2 Analysis of domestic finance-related data gathered through the country-level case studies' data collection. 6.3 Interviews with Secretariat domestic finance specialists and qualitative content analysis. 6.4 Interviews with ITAP experts on domestic finance and qualitative content analysis.
7. How well does GPE design and mobilize assets beyond national borders to support	7.1 How have GPE's cross-national mechanisms (particularly EOL) been used to identify system bottlenecks and advocate for and formulate reforms	7.1 EOL-funded advocacy has led to/supported greater visibility for domestic finance issues at the country level or led to/supported domestic finance-related policy reforms.	7.1 Desk review of EOL grants in case study sample countries: EOL evaluation.

Leading question	Subquestion	Indicators	Data sources and analysis
improvements in domestic finance in partner countries?	with respect to domestic finance in partner countries? 7.2 To what extent has GPE's participation in global forums/platforms, such as the Transforming Education Summit, helped to advance progress on the volume, equity and efficiency of domestic finance (e.g., by aligning actors behind GPE's strategy, advocating for greater attention to domestic finance issues and building a common understanding of domestic finance issues)?	7.2 GPE's participation in global/regional forums/platforms has led to greater visibility for domestic finance issues/alignment of other actors behind GPE's approach to domestic finance etc.	7.2 Interviews with EOL Grant Agent and Secretariat EOL lead and qualitative content analysis. 7.3 Desk review of outcome documents from global forums/platforms in which GPE has participated. 7.4. Interviews with Secretariat domestic finance specialists and Education Policy and Learning lead and qualitative content analysis.
Baseline (gender equality thematic case study)			
8. What is the status of partner countries with respect to gender equality in, and through, education? (Refer to GPE guidance for a definition of gender equality in, and through, education).	8.1 What is the status of partner countries with respect to gender equality in and through education? 8.2 To what extent, and how, is the role of gender equality in, and through education, being understood and conceptualized in different country contexts? To what extent is gender hardwiring in policies and programs understood? 8.3 What is the level of government and partner capacity to identify and address system bottlenecks with respect to gender equality and to hardwire gender equality considerations in all policies and programs?	8.1 Indicators on gender equality in and through education in gender analyses conducted by countries; gender-related guiding considerations from the enabling factors self-assessments (not limited to those on gender-responsive planning and monitoring); other standard indicators of gender equality. 8.2 Extent to which gender equality in and through education, and the key underlying concepts, are understood and considered important by key education stakeholders. 8.3 Extent to which country-level stakeholders have/report having sufficient capacity and access to data to carry out diagnostics on gender equality (including the enabling factors self-assessment) and develop solutions to address the gaps identified.	8.1 Desk review of gender diagnostics (including gender dimensions of enabling factors self-assessments, ITAP assessments and country responses to ITAP assessments); collection and analysis of gender equality-related data from other standard sources (UIS, education sector analyses etc.). 8.2 Analysis of gender equality-related data gathered through the country-level case studies' data collection. 8.3 Interviews with Secretariat gender specialists and qualitative content analysis.

Leading question	Subquestion	Indicators	Data sources and analysis
9. What, if any, policy actions and prioritized reforms are partner countries choosing to undertake with respect to gender equality? Are the prioritized reforms and policy actions likely to be transformative and are partner countries well-positioned to successfully implement them?	9.1 What gender equality-related policy actions and/or prioritized reforms are partner countries undertaking? To what extent are they aligned with gaps identified through gender analyses? 9.2 To what extent does the successful implementation of the policy actions/prioritized reforms have potential to address gender equality gaps? 9.3 Have gender equality dimensions been adequately hardwired in prioritized reforms that are not specifically gender equality-focused?	9.1 The gender equality-related policy actions and prioritized reforms in the partnership compacts are evidence-based (including being aligned with gaps identified through diagnostics), appropriately designed and sufficiently ambitious to address the challenges identified. 9.2 Prioritized reforms in partnership compacts that are not specifically gender equality-focused, identify and address the gender barriers related to the reform that girls and boys face.	9.1 Desk review of gender diagnostics (including enabling factors self-assessments, ITAP assessments and country responses to ITAP assessments) and gender-related policy actions and prioritized reforms in partnership compacts; collection and analysis of gender-related data and policy documents from other standard sources (UIS, education sector analyses etc. and government policy documents). 9.2 Analysis of gender-related data gathered through the country-level case studies' data collection. 9.3 Interviews with Secretariat gender specialists and qualitative content analysis.
10. What support, with respect to gender equality, is GPE providing to partner countries to design and align resources to their chosen priority reforms? How relevant, efficient, effective and coherent is GPE support in helping countries to identify transformative reforms in gender equality and align resources behind them?	10.1 To what extent did the enabling factors self-assessments and compact dialogue (including country self-assessment and ITAP assessment) support the effective identification of challenges in gender equality (this is not limited to the gender-responsive planning and monitoring enabling factors because gender equality dimensions are included in all enabling factors)? What aspects/features (process steps, methodology, guidance etc.) of the process were most effective? How efficient was the process? 10.2 To what extent has GPE's allocation mechanism (minimum allocation + top-up) served as an incentive to address system	10.1 Extent to which enabling factors self-assessments identify the key gender equality-related system bottlenecks; are evidence-based, of quality and inform policy dialogue; feasible given country capacity; country-level stakeholders believe that the value added by the domestic finance self-assessments justifies the cost of preparing them. 10.2 Extent to which ITAP assessments of the gender equality dimensions of the enabling factors are robust and evidence-based; fulfill their role of providing independent, technical assessments of country status with respect to the enabling factors areas; consistent across countries; inform country-level policy dialogue. 10.3 Extent to which the top-up portion of the STG served as an incentive to address gender equality-related gaps in the enabling factors.	10.1 Desk review of gender equality-related dimensions of the enabling factors self-assessment documents based on criteria corresponding to the indicators/success criteria (including ITAP assessments and country responses to ITAP assessments) and gender equality-related prioritized reforms and policy actions in partnership compacts; collection and analysis of gender equality-related data and policy documents from other standard sources (UIS, education sector analyses etc. and government policy documents). 10.2 Analysis of gender equality-related data gathered through the country-level case studies' data collection.

Leading question	Subquestion	Indicators	Data sources and analysis
	bottlenecks in the gender equality dimensions of the enabling factors? 10.3 To what extent did the partnership compact process shift the dialogue among education partners toward gender equality in, and through, education? To what extent did the process align partners behind the chosen policy actions and prioritized reforms with respect to gender equality? 10.4 To what extent did the SCG support partner country capacity to address gender equality? Was the capacity support provided aligned with country needs? 10.5 To what extent did the GEA serve as an additional incentive to hardwire gender equality dimensions in countries' prioritized reforms (in eligible countries)?	10.4 Extent to which partnership compact processes supported/led to greater stakeholder alignment behind policy actions and prioritized reforms focused on gender equality and behind interventions to address gender-related barriers to education faced by boys and girls that are embedded in other reforms. 10.5 The enabling factors self-assessment and compact processes strengthened the capacity of country level stakeholders to diagnose gender equality-related system bottlenecks, identify corresponding reforms and policy actions and align stakeholders behind the reforms and policy actions identified. 10.6 Gender equality-related activities financed through the SCG are aligned to country needs and are likely to strengthen country capacity. 10.7 The GEA served as an additional incentive to hardwire gender equality dimensions in countries' prioritized reforms and policy actions.	10.3 Interviews with Secretariat gender specialists and qualitative content analysis. 10.4 Interviews with ITAP gender experts and qualitative content analysis.
11. How well does GPE design and mobilize assets beyond national borders to support improvements in gender equality in partner countries?	11.1 How have GPE's cross-national mechanisms (particularly KIX and EOL) been used to identify challenges and advocate for and formulate reforms with respect to gender equality in partner countries? 11.2 To what extent has GPE's participation in global forums/platforms, such as the Transforming Education Summit, helped to advance progress on gender equality (e.g., by aligning actors behind	11.1 EOL-funded advocacy has led to/supported greater visibility for gender equality issues at the country level or led to/supported gender equality-related policy reforms. 11.2 KIX-funded knowledge generation and exchange has led to improved policies and/or country capacity to address gender equality-related issues. 11.3 GPE's participation in global forums/platforms has led to greater visibility for gender equality issues/alignment of other actors behind GPE's approach to gender equality etc.	11.1 Desk review of EOL and KIX grants in sampled countries; EOL and KIX evaluation. 11.2 Interviews with EOL and KIX Grant Agents and Secretariat EOL and KIX leads and qualitative content analysis. 11.3 Desk review of outcome documents from global forums/platforms in which GPE has participated. 11.4 Interviews with Secretariat gender specialists and qualitative content analysis.

Leading question	Subquestion	Indicators	Data sources and analysis
	GPE's strategy, advocating for greater attention to gender equality issues and building a common understanding of gender equality issues)?		

5   Methodology

5.1   Methodological Approach Overview

A **mixed-methods** approach has been selected for this evaluation. Mixed-methods research entails a purposive combination of qualitative and quantitative methods for data collection, data analysis and interpretation of the evidence, which ensures robustness and minimizes bias. By purposively combining techniques for data collection, analysis and interpretation, we can select the best method to answer each evaluation question and leave open the opportunity to adapt our approaches as the evaluation both evolves and progresses over time.

Figure 7 demonstrates the overall and iterative methodological process that we will undertake to operationalize this evaluation, and which is described in the remainder of this section. This process can be considered as both the overarching steps involved in the longitudinal country-level case studies (steps 1–12, which cover the iterative activities for each of the baseline, midline and endline phases) in which the steps build on each other each phase, as well as the steps (steps 13–16) that are repeated at each phase of the implementation of our evaluation (e.g., Phases 1, 2 and 3) in order to move from the country-level case studies to our wider synthesis report. It is important to note that the diagram below implies a linear sequencing of events, which is not necessarily the case, as some steps are not sequentially dependent and therefore may be concurrent (such as the adapting the study steps and step 13 on global-level key stakeholder interviews, which may run concurrently to country-level case study data collection and analysis).
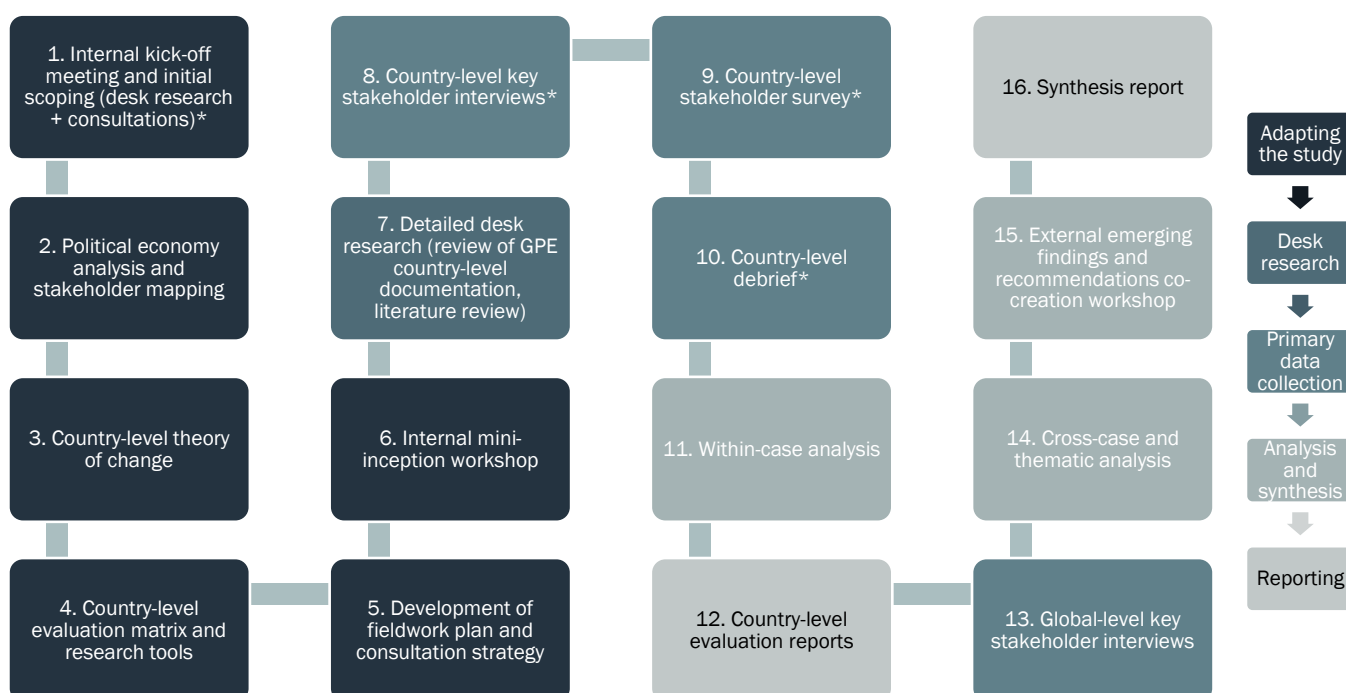

Figure 7. Overview of the iterative process for adapting the study, data collection and analysis. Steps 1–12 apply to our country-level case studies and should be considered as an iterative process applying to the baseline, midline and endline. Steps 13–16 are repeated for each phase to support the synthesis process, culminating in our synthesis report.

* Indicates touch points with in-country stakeholders.

For the country-level case studies, the baseline begins with an extensive period to adapt the study design (steps 1–6), before proceeding into the detailed desk research and primary data collection (steps 7–10) and then within-case analysis (step 11), in order to produce each country-level case study report (step 12). The same process is iterated for midline and endline, in which we repeat each step

involved, although some of the steps (such as steps 2, 3 and 5) will likely only require refreshing and updating outputs from the baseline.

To support the synthesis process, in each phase we will conduct a set of high-level global-level interviews (step 13) to support both the thematic case studies and our cross-case analysis. These will not necessarily follow the country-level case study process, but may take place concurrently or in response to country-level case study findings. To support our cross-case and thematic analysis, we will hold emerging findings and recommendations co-creation workshops with stakeholders from the GPE Secretariat (step 15) ahead of finalizing our synthesis report (step 16). These steps are repeated in Phases 1, 2 and 3, and apply whether looking at country-level case studies at baseline, midline or endline.

For this section of the report, we describe each step in this process. The section is organized in terms of types of activities (implied in the colors in figure 7), rather than by the numbered sequence. Each subsection will describe the methods and approaches that apply for both the country-level case studies as well as for our synthesis process, as relevant.

We will also explain how these processes might differ across different phases of the evaluation. Given the phased roll-out and progressive design of our evaluation, there will be adaptations required to this process depending on the phase. This includes whether activities will take place in-country, which may differ across phases. Where possible, we have described how this process may differ across each phase. Ahead of the start of Phases 2 and 3, we will review and update this process, which we will describe in a concept note. Finally, this section also presents some of the limitations of our methodology.

Box 2. Adapting the evaluation and methodology in different phases

The process outlined in section 5.2 primarily relates to the first instance of each country-level case study (e.g., the baseline), which will be conducted in Phase 1 for the eight Wave 1 countries and Phase 2 for seven Wave 2 countries.

In subsequent phases for both waves of studies, we will undertake an abbreviated version of this adaptation process (e.g., at midline or endline). This allows us to ensure that the country-level case study design progresses to the next set of questions, is continuously relevant to country context over time and incorporates findings from the previous phase to further develop the country-level theories of change. We will update our country-level political economy analysis (PEA) and theory of change as a result of the findings from the previous wave, but will not re-conduct a country-level theory of change workshop. Our country-level evaluation matrix and research tools will be updated to reflect the new evaluation questions which apply to this new phase, and we will update our fieldwork plan and stakeholder consultation strategy accordingly.

Finally, our staggered roll-out to the evaluation allows us to both stagger the roll-out of country-level case studies within phases as well as across phases. In Phase 1, we will stagger the roll-out of our Wave 1 country-level case studies by one month, first implementing four and then rolling out an additional four. This allows us to learn and iterate as we proceed, making adaptations to protocols, processes or data collection tools as required.

We are able to reflect on our methodology and processes at the end of Phase 1 and translate these learnings into the roll-out of the Wave 2 country-level case studies in the following phase. At the end of Phases 1 and 2, we will take the opportunity to reflect on whether any changes are required to the evaluation and case study designs, which will be summarized in a concept note designed to capture and describe proposed changes to the evaluation, as well as look forward to determine how the evaluation should progress to assess the next stage.

5.2 Adapting the Evaluation to Country Context

It is critical that the overall evaluation approach is adapted at the start of each country-level case study so that it fully aligns with the particular needs of each country context. Adaptation within the overall framework will ensure that the studies maintain consistency and that each one is linked to the overall

evaluation matrix. Therefore, at the start of the baseline for each study, we tailor our overall evaluation approach to each country-level case study. Prior to the midline/endline, we will undertake an abbreviated version of the adaptation approach laid out here to ensure that the case study design is still relevant and has been further adapted where necessary.

This adaptation process will be led by the country-level case study teams – teams of evaluators with relevant expertise brought together to conduct the country-level case studies from start to finish across all three evaluation phases – under the overall direction and guidance of the core evaluation team. More information on the structure of our team is found in section 8.6.

5.2.1 Internal Kick-off Meeting and Initial Scoping (Initial Desk Research and Consultations)

To ensure that the overall evaluation approach is adapted appropriately, at the start of each baseline each country-level case study team will begin with an internal **kick-off meeting** with the core evaluation team to ensure alignment and understanding of the GPE strategic framework and operating model, evaluation theory of change, generic evaluation matrix and generic data collection tools and to discuss broad plans for conducting the country-level case studies.

Following the kick-off meeting, country-level case study teams will conduct an **initial scoping** of country-specific documentation related to the in-country compact development process and the selected priority reform as well as a consultation session with relevant stakeholders from the GPE Secretariat, namely the Country Team Lead (CTL) for the partner country in question. This allows each country-level case study team to develop an initial understanding of the stakeholder landscape. Country-level case study teams can also discuss with the CTL how to tailor the approach required for the country, including any specific requirements for stakeholder communications or data collection protocols. If any key documentation or information required is missing, country teams will use this session to fill information gaps and seek additional documentation from the CTL. Finally, the CTL will be introduced to the timelines and varied roles and responsibilities, including theirs, for the country-level case study throughout the three phases.

In some cases, it may be necessary to seek the support of CTL to broker preliminary consultations with key country-level stakeholders if clarifications are required on the country-level theory of change. This will be determined as part of the initial desk research. For further information on this, see below on the development of country-level theories of change.

Subsequent phases will only involve a short kick-off meeting to introduce to the country-level case study teams the revised evaluation questions that apply to the next phase and provide any updates to design as necessary, and short updates to the CTL on the relevant activities and timelines.

5.2.2 Political Economy Analysis and Stakeholder Mapping

GPE partner countries each have their own political and economic conditions, which will significantly affect the compact development process, its operationalization and wider education transformation. Furthermore, the success of development process and leveraging priority reforms for system transformation is contingent on actor alignment. Therefore, country-level case studies need to carefully consider contextualized understanding of the different and possibly divergent dynamics between and interests of key systems actors and the institutions they represent.

Country-level case study teams will use desk-based review to form our initial **political economy analysis** (PEA). Sources for this include the information that has been compiled by partner countries to support their compact development process as well as sector analyses conducted by the World Bank and other institutions. The PEA serves as an essential starting point for understanding the state of play of the education system and policy in the country, the wider sociodemographic environment and broad

political context. Our approach to PEA is in line with the PEA practice explored and researched recently by the Research in Improving Systems of Education (RISE) team.[51]

The main areas of focus for PEA include the following:

- Education system governance structure, including level of centralization/decentralization; roles of national, regional and local authorities; mapping of relevant government ministries, agencies and other relevant institutions (including non-governmental institutions); and mapping (to the extent feasible and relevant) the relationships between these, and resourcing/capacity and power dynamics relating to gender equality actors. This will particularly investigate dynamics that could facilitate or hinder the implementation of a system-transforming reform.

- Education system objectives, as stated in key policy documents and other official government materials.

- Main recent reforms, strategies and needs analyses; understanding of reform cycle and current timing in reform cycle; understanding of political cycle and current timing in political cycle, and the degree to which they overlap with GPE priority areas (in particular gender equality and domestic finance).

- Lessons learned from implementation of reforms in the past (including constraints to their success, failure of technical design, failure of capacity and inadequate resources) in order to assess whether the priority reforms in compact are likely to address those constraints.

- Major donor programs in-country and their dynamics in relation to government reforms; donor coordination and alignment mechanisms.

- Current sociopolitical state of the partner country (e.g., existence of conflict, crises and political instability).

- Current state of education financing, trends over time and understanding of education financing relative to other/similar countries.

The results of the PEA will be summarized and included as an annex to our baseline country-level case study reports. For cross-comparison purposes, each PEA annex will be prepared following a standard template.

A critical component of our PEA and our country-level case studies is the development of a **country-level stakeholder map**. This serves to inform the identification of the relevant stakeholders to participate in our stakeholder survey and interviews, as well as to support our PEA by examining and analyzing the dynamics between these actors. Building on the initial scoping and consultation process, each country-level case study team will further examine documentation (including the terms of reference for local education groups) to identify the list of stakeholders involved in design and implementation of the partnership compact (including, but not always exhaustively the members of the local education group) as well as the education system more broadly.[52]

It will be particularly important to compare the wider landscape of actors in the education system with those involved in the compact development process to consider the inclusivity of the processes involved in the compact development and the implementation of the priority reform in later phases. We will establish an indication of the strength that engagement stakeholders have within the compact and their general degree of interest and influence. We will also consider which voices may not have been represented in the compact development process, including those not directly engaged with GPE but

---

[51] See for example: Levy, B. *How Political Contexts Influence Education Systems: Patterns, Constraints, Entry Points*, RISE Working Paper 22/122, 2022, available online at: https://riseprogramme.org/sites/default/files/2022-12/How_Political_Contexts_Influence_Education_Systems.pdf

[52] This includes the stakeholders involved in the local education group, such as the Ministry of Education representatives, Coordinating Agencies, Grant Agents, international organizations including donor and multi-lateral organizations, and representatives from civil society and teacher organizations. This will importantly also include those involved in education service delivery with a focus on the priority reform, such as at various levels of the government (district, local government etc.), different departments within the ministry of education, representatives from Teacher Training Colleges, Teacher Service Commissions, and where relevant, school leaders and teachers.

working within policy environments directly associated with education, in order to bring in the systems-level approach and to bring an additional perspective and "critical voice" into the discourse.

We resourced our country-level case studies so that the bulk of this work will happen as part of the baseline for each study. This process will not be repeated in full in subsequent phases, but the PEA and stakeholder map will be subsequently updated in each phase, building on the findings from primary data collection and any new secondary evidence.

5.2.3 Development and Refinement of a Country-Level Theory of Change

In conjunction with the initial desk review and PEA, country-level case study teams will each develop a country-level theory of change. The starting point for this will be the theory of change developed as part of the country's partnership compact and the overarching evaluation theory of change, and will articulate and describe the intended transformation pathways at the country level. We anticipate that the earlier phase depicted in the theory of change (e.g., on the compact development) is likely to be similar for each country, whereas the segments of the theory of change on compact implementation, and of the transformative reform and objectives, will be more varied as each country will have its own unique transformative reform and therefore a unique change pathway.

If the theory of change articulated in the partnership compact does not contain a sufficient level of detail in order for us to adapt and align it to the evaluation's wider theory of change, it may be necessary to have further consultation on the theory of change. In this situation, we may need to validate our understanding and seek further details to inform the remainder of design of the country-level case study, including the refinement of the country-level evaluation matrix and tools. Country-level case study teams will endeavor to identify this challenge ahead of the initial consultation session with the CTL so that the CTL may act as another potential source of information. If this does not provide further information, the country-level theory of change will instead use the primary data collection phase to consult with stakeholders to flush out the theory of change.

This process will not be repeated in full in subsequent phases as this theory of change will serve to guide the country-level case study throughout its phases and may be refined to reflect an updated understanding or evolution of the pathways.

5.2.4 Development and Refinement of a Country-Level Evaluation Matrix and Research Tools

Following the development of the country-level theory of change, each country-level case study team will develop a country-level evaluation matrix which builds on the evaluation's overall matrix (and evaluation questions relevant for the phase) and reflects the country's individual theory of change. We will also take into account the ways in which each country-level case study should be adapted to address the cross-cutting themes of gender equality and domestic financing, where relevant. The country-level evaluation matrix will also inform the corresponding adaptation of research tools to the country-level and for different stakeholder types. The research tools will be quality assured internally following the mini inception workshop (see below) before data collection begins.

This process will be repeated for each phase as new evaluation questions are introduced to the study to reflect the progressive nature of the evaluation and therefore the country-level case studies (as discussed in sections 3.2 and 4.3). These will require updates to the questions in each of the data collection tools.

5.2.5 Development of Fieldwork Plan and Drafting a Consultation Strategy

Each country-level case study team will develop a fieldwork plan and stakeholder consultation strategy. These will set out the planned activities for the phase and allow the core evaluation team and the Evaluation Manager to assess any additional risks posed by data collection as well as to monitor and support the process of data collection and ensure that the evaluation's guiding principles and research ethics are upheld throughout the data collection process.

The stakeholder map (developed in the PEA stage) will be operationalized through a consultation plan, which will outline which stakeholders are to be consulted during data collection, through which modality

(including data collection and other forms of engagement), throughout the course of the country-level case study.

These planning processes will also help to ensure that data collection is conducted in an efficient manner, in order to reduce the probability of research fatigue.

5.2.6 Internal Country-Level Mini Inception Workshop

The adaptation process will culminate in an internal country-level mini inception workshop, in which country-level case study teams will present the results of their desk-based research and their adapted theory of change, evaluation matrix and tools back to the core management and evaluation teams.

This is an important quality assurance (QA) step to ensure that country-level case studies have been adequately and appropriately adapted by country-level case study teams and provides the core management and evaluation teams with feedback on how effectively the overall evaluation approach can be adapted at the country level.

As studies are also staggered within each phase, the workshop also provides an opportunity to confirm which processes may need to be adapted ahead of roll-out of the set of studies in the wave. Following the workshop, research tools will be additionally quality assured by internal core management team members and then shared with the GPE Secretariat's R&P team for further review (if needed) before data collection begins.

5.3 Desk Research

Our evaluation aims to make use of a comprehensive desk review to support our country-level case studies. This stage builds on the initial scoping and PEA development (indicated in section 5.2). The earlier stage was intended to inform the design of the country-level case studies by identifying political economy factors and focused on understanding the compact development process, as well as information and data that can help to refine and focus the data collection tools.

At each phase, it is anticipated that we will require further desk research in order to collect further documentation and secondary data related to the key evaluation questions for each phase. As part of this fuller desk research process, we will make use of the documents, data and evidence compiled by partner countries as part of the compact development process (including but not limited to those shared with the GPE Secretariat), which not only includes information on the compact development process, but the wider context of education reform. We will also supplement this with further review of country-level academic and gray literature, including documentation on other programs and initiatives, which have similar activities or alternative delivery models in the data collection phase. Finally, as the evaluation proceeds, we will also gather and collate data collected by the GPE Secretariat and country-level stakeholders related to the implementation of priority reforms and GPE grants.

In table 3, we outline the key purposes of our comprehensive desk review and the associated sources of information, which we have determined through our initial document mapping process completed during this inception phase. We will use data capture templates and make use of a common coding framework to support document coding (see section 5.5.1 for more information), which will support in the aggregation of data and ensuring that our desk reviews remain structured and comparable, supporting our wider data synthesis efforts.

Table 3. Document review purpose and corresponding documents

Purpose	Types of documents
Background understanding country's policy needs and policy landscape as well as previous education reform	Documents related to previous country education sector plans and supporting policy documents
	Country analytics document
	Documents compiled as part of the enabling factors analysis (including to support the "data and evidence" enabling factors)

efforts to inform our PEA (baseline)	Literature review (country-level academic and gray literature) with relevance to the GPE model, the pathways and selected priority area
	Documentation on other programs and similar initiatives
Background understanding of country's previous engagement with GPE to inform our PEA (baseline)	Previous prospective or summative country-level evaluations (where available)
	Engagement memo and country analytics
	Previous GPE grant program documents and completion reports
To identify main actors and key stakeholders for participation in the evaluation (stakeholder mapping and consultation strategy) (baseline)	Local education group membership lists and local education group Terms of Reference
	Documents related to coordination factor as part of the enabling factors analysis
	Literature review (country-level academic and gray literature) with relevance to the GPE model, the pathways and selected priority area
To provide information on the country's compact development process (including relevance and effectiveness of the process) (baseline)	Country analytics documents, engagement memos and enabling factors analysis
	ITAP report and local education group comments
	Country compacts and related documents
	Board decision and strategic parameters documentation
To explore the effectiveness of the implementation of the priority reforms (midline, endline)	GPE grant-level documentation (including any monitoring data collected as part of the grant) and data collected to support the GPE results framework
	Documentation and data generated at the country-level (by government, civil society, international actors etc.) related to the compact and priority reform implementation and for tracking the priority reform's intended outcomes

5.4 Primary Data Collection

Primary data collection forms a key part of our country-level case studies across all phases. Our primary data collection will utilize two main methods for data collection: key stakeholder interviews (at the country and global levels) and country-level stakeholder surveys. We will use both in-person and remote data collection.

For country-level case studies, in-person data collection (both country-level key stakeholder interviews and stakeholder survey) will center around five-day field visits involving our country teams used at baseline for both Waves 1 and 2, and during endline (Wave 1) and midline (Wave 2).[53] As our country-level case studies are longitudinal, the opportunity for iterative but efficient contact points with key stakeholders will allow our country-level case study teams to build relationships and trust with research participants and make planning for remote data collection and for the second visit more straightforward. Our country-level primary data collection will conclude with data collection debriefs with data collection participants in order to validate and provide feedback on the data collected.

We will supplement our in-country data collection with remote global-level key stakeholder interviews, which support our thematic case studies and synthesis process. These are summarized in table 4 and expanded upon in the next sections.

---

[53] Visits will occur twice for each country: in the first instance of the country-level case study (e.g. in Phase 1 for eight country-level case studies and in Phase 2 for the additional seven country-level case studies) and in Phase 3 (for all 15 country-level case studies).

Table 4. Summary of primary data collection to be used in all phases

Data collection method	Rationale	Suggested sample
Country-level key stakeholder interviews	To explore the process of the operating model roll-out from the country-level perspective (baseline) and views on the context and implementation of priority reforms, including as related to the transformative potential of the reform and on thematic areas (all phases as relevant)	~24 interviews per country per phase Ministry of education, focal point, Grant Agent, Coordinating Agency representatives Other local education group members including from civil society organizations, teacher organizations, international organizations and donor groups Stakeholders outside of local education groups or education ministries who play a role in education policy and service delivery Key national education experts
Country-level stakeholder surveys	To closely examine the extent of education stakeholder alignment and inclusion and changes over time	~24 respondents per country per phase (as with country-level key stakeholder interviews)
Global-level key stakeholder interviews	To understand process aspects of the operating model roll-out (baseline) and ongoing GPE support (both country-level and thematic) (all phases)	~15 interviews per phase GPE Secretariat CTL (and if relevant, other team members supporting country-level operations including education specialists and grant operations officers) Grant agents global or regional focal points GPE Board members GPE Secretariat staff (including from CEP, FGO and R&P) Members of ITAP Key education experts

It should be noted that for all primary data collection, we follow GPE's guidelines for evaluations to be conducted with adherence to the highest ethical principles which include data minimization, privacy and confidentiality, informed and voluntary consent and respect for cultural sensitivities and human rights. This is described in section 3.3 and more information is found in **Annex 8** on research protocols.

5.4.1 Country-Level Key Stakeholder Interviews

In-depth interviews with key stakeholders are the primary means of collecting rich qualitative data to robustly answer the evaluation questions, create assessments against the criteria for system transformation (as outlined in section 4.1) and to support our investigation of the theory of change and assumptions, including the realist approach questions of "why, why not" and "how."

As outlined in section 5.2 on the development of a country-level evaluation matrix, at each phase the key stakeholder interviews will focus on different questions that will reflect the progressive roll-out of the compact and priority reform, as well as the relevant thematic areas.

The focus of our country-level fieldwork period will be country-level stakeholder interviews. At the baseline for each country-level case study this will mainly involve face-to-face interviews and will be conducted in the relevant local language. In-person data collection will be repeated in Phase 3 (e.g.,

endline for Wave 1 countries and midline for Wave 2 countries). During the midline for Wave 1 studies, this will only involve remote interviews. We anticipate that in total, for each country-level case study, we will interview approximately 24 stakeholders per phase.[54]

In each country-level case study, our country-level stakeholder interviews will be guided by the stakeholder consultation strategy developed in the country-level case study planning phase (see section 5.2) to set up interviews, although we may use a snowball approach to secure further introductions once interviews are underway. We have already highlighted the importance of engaging with diverse voices in-country. We will also include policy makers outside education ministries who are not directly engaged with the partnership compact but are engaged with education at a systems level (for example those with a portfolio for finance and planning, gender etc.).

To broaden the range of interviewees, where relevant, country-level case study teams may opt to conduct group interviews with 6–8 participants per group. These group discussions will bring together groups of similar types of stakeholders (such as members of civil society) and allow us to capture a greater number of perspectives.

To ensure that the interviews do not overburden respondents, all interviews will be limited to 45–60 minutes each, while group interviews may be extended to 90 minutes with prior agreement from participants. Interviews will be conducted in the relevant operating language of the country with the support of translators if required.

5.4.2 Country-Level Stakeholder Surveys

We will conduct a short survey as part of each country-level case study to collect specific information on the dynamics of the relationships between in-country stakeholders. The survey will prioritize collecting data from members of the local education group (although if viable, we may also collect data to explore relationships between actors within local education group and outside). Survey questions will be designed to assess the strength of the relationships between the stakeholders, including levels of trust, collaboration, influence and engagement and serves the basis for our social network analysis (SNA – see section 5.5.1). The survey will be repeated in each country-level case study across all phases in order to track any changes in the partnership dynamics over time.

We understand from the initial consultations with the GPE Secretariat that there have been challenges in attaining high response rates from local education group members from previous survey efforts. We designed our survey instrument and survey protocols with this challenge in mind and outline below the strategies and features we will include as part of our survey to maximize response rates. We will also assess the method's effectiveness after Phase 1 to ensure that we are using the best method to collect these data.

Our survey will be designed to be short (approximately 15 minutes), using closed-ended questions. To simplify the survey process, we propose to initially administer the survey with the same respondents as country-level key stakeholder interviews, either as part of the interview itself or as a follow-up. These respondents will have been selected on the basis of our stakeholder mapping and will include a diversity of representatives from across the local education group as well as outside of the group.

The survey will be designed using an online survey platform (such as Alchemer[55]), and will be administered face-to-face (with the interviewer acting as the enumerator) as well as through an online self-administered questionnaire. This allows us the flexibility to open the survey up to a greater number of respondents (such as beyond our interviewees). This will be on the basis of the validation of the survey effectiveness (in terms of response) and using a snowball approach following our stakeholder map.

If we opt to open the survey to a larger number of respondents, the questionnaire will be designed to include routing (to tailor questions to the type of respondent) and will amend our stakeholder consultation strategy to include country-specific survey protocols to guide survey engagement. These

---

[54] This figure is based on the amount of time allocated for country-level fieldwork, which is five days. Up to three country-level case study team members may be conducting interviews simultaneously, but it is also anticipated that interviews will require multiple team members.
[55] https://www.alchemer.com/

protocols will set out basic parameters for survey conduct to ensure clear communications about the survey purpose and process, as well as to carefully manage communication on the survey beyond those interviewed.

5.4.3 Country-Level Data Collection Debriefs

We will conduct country-level debriefs at the end of the data collection stage of each country-level case study and ahead of further analysis. Where in-country fieldwork has been conducted, these workshops will be held in-country. The workshops serve to present our main fieldwork findings back to our respondents to support their validation and allow opportunity for clarification or follow-up. The main audience of these debriefs will be stakeholders interviewed during fieldwork, but additional participants may be identified as part of the stakeholder mapping process and in consultation with CTLs or Coordinating Agencies (and other stakeholders involved in initial country-level consultations).

Country-level data collection debriefs can also serve an important role in engaging country-level stakeholders as part of the learning process. and support our approach to developmental evaluation. This is discussed in greater detail in section 6.

5.4.4 Global-Level Key Stakeholder Interviews

Global stakeholder interviews will be conducted virtually with GPE Secretariat staff (including CEP and particularly GPE priority area leads and domestic financing/gender specialists), members of ITAP, GPE Board members and key experts. These interviews will be conducted by the core evaluation team and serve to supplement and follow the country-level data collection efforts.

Global stakeholder interviews will initially focus on exploring aspects of the design of the operating model and its centralized roll-out as well as ongoing post-compact GPE support and model evolution in subsequent phases. Global interviews will also explore thematic areas, such as on how thematic priority areas are conceptualized and promoted through operating model roll-out. We will apply a flexible number for these but anticipate up to 15 interviews per phase based on our initial count of the number of stakeholders relevant for data collection at baseline/Phase 1.

5.4.5 Data Collation, Storage and Systematization

Both primary and secondary data will be stored on Triple Line's bespoke and secure SharePoint site. More information on our data security processes can be found in section 8.4 and **Annex 8**.

Survey responses will be collected using an online platform (such as Alchemer[56]), which reduces the need for manual data entry and possible data entry error. Where surveys are administered face-to-face, the researcher will complete the online questionnaire as part of the process.

The raw data will be securely and digitally archived and backed up at a second location, so they are available for retrieval if necessary for QA purposes. The raw data will then be de-identified (with identification information stored using a separate password-protected key) and cleaned (to ensure that data values fall within and conform to an expected range) prior to analysis.

All interviews and focus groups will be audio recorded (with respondent's consent)[57] and then transcribed using structured templates, allowing responses to be recorded against questions in the topic guide. Recordings will be stored separately to write-ups that will be de-identified to protect the confidentiality of the respondents. The write-ups will be reviewed for any missing information (e.g., where the audio quality is poor in the case of online interviews, or there is insufficient time to cover all questions) or where the meaning behind a response given is not clear. Where interview write-ups lack important information, or certain responses require further clarification, country-level case study teams will follow up with respondents to request clarifications.

---

[56] https://www.alchemer.com/

[57] As professional researchers and evaluators, we pay particular attention to treatment of data and compliance with GDPR as is outlined our privacy policy. The consent process will be outlined as part of the research tools. Further information on this can be found in **Annex 12** on Research Protocols.

5.5 Analytical Methods

There will be a series of analytical techniques used to analyze the desk review and primary data collected during the country-level case studies. In this section, we present the different techniques used for both **within-case analysis** (e.g., analysis within our country-level case studies) and as part of our **cross-case analysis** (which looks across our country-level and thematic case studies, feeding into our synthesis report).

Within-case analysis will be conducted by each country-level case study team whereas our cross-case synthesis will be conducted by our core evaluation team.

As the evaluation progresses across phases and explores different aspects of system transformation (e.g., the potential for system transformation, early implementation of the transformative priority reform and progress toward/achievements of system transformation and associated impact), different analytical methods may apply. Therefore, this section describes the analytical methods that will be used throughout each of the phases, where relevant specifying when the analytical method will be applied.

5.5.1 Within-Case Analysis

Qualitative Content Analysis

Qualitative content analysis is a method used to interpret meaning from the content of textual data, identify themes and topics and assess the extent of their emphasis across multiple pieces of evidence. This implies using pre-defined codes or labels based on the research/evaluation questions and then analyzing how strongly those elements appear in the collected evidence.

We will perform content analysis across a range of qualitative data, as follows:

- **Primary data:** country-level stakeholder interviews or group interviews
- **Secondary qualitative data:** GPE documentation, GPE Results Framework, country-level documents and data.

A separate qualitative content analysis will be performed for each of the above data sources and analyzed on the basis of the indicators set out in our evaluation matrix, in our theory of change assumptions, and on the basis of the system transformation criteria set out in section 4.1.

Our evaluation will likely produce a large volume of qualitative data, which will require careful organization and centralized management in order to structure the analysis processes. We will use the specialist computer-assisted qualitative data analysis software ATLAS.ti[58] to store and organize the data systematically and this will enable the data's rapid retrieval and structured analysis.

All relevant data for the evaluation (including primary and secondary data) will be stored in ATLAS.ti. Data will be organized using "codes," which will primarily be pre-determined (e.g., deductive coding) and will cover a variety of categories (types of documents and of respondents), concepts (related to the evaluation matrix) or themes. We will use a unified coding framework to ensure that all data are consistently organized into codes, both across country studies and over time, allowing us to maximize the data analyzed and for quick retrieval. Our coding framework will be regularly reviewed, allowing us to incorporate new ideas or unexpected answers (e.g., developing new "inductive" codes) as part of our analysis. We will follow a standard practice based on the following principles in the development of our coding framework: balance between the ability to cover as many responses as possible and usefulness for the analysis, the unambiguity of the code and the ability to capture both positive and negative feedback and ideas.

Once data are assigned to their relevant codes, we can quickly retrieve data by running "data queries." This also allows us to produce quasi-quantitative overviews of the frequency distribution of metrics of interest (e.g., engagement of leadership, political leverage in place and engagement of civil society) and to use visualization techniques, such as graphs and charts. We will also be able to aggregate or compare data across studies and over time as part of the cross-country synthesis process. This allows for both a

---

"top-down" analysis of results, as well as allowing for a "bottom-up" emergence of new theories to arise from the data (a more inductive approach).

Quantitative Analysis

Quantitative analysis will be performed across the following data:

- survey with country-level stakeholders (closed questions)
- GPE 2025 Results Framework (when the data are of quantitative nature)
- national statistical data sourced at the country level and the international level (e.g., World Bank databases, UIS statistics by UNESCO or partner countries' ministries of education).

Separate quantitative analysis will be performed for each of the above data sources for each country.

Our primary means of analysis will be based on **descriptive statistics**. This may be complemented by advanced statistical analysis to provide further insights into the differences and similarities between the different categories (e.g., stakeholder groups in the case of the stakeholder survey) and to test whether those observed differences are statistically significant.

Social Network Analysis

The SNA is an analytical technique allowing the study of **social relationships within a network**. It provides evidence on which members of the network tend to have more influence and create more relationships than others and helps create a picture of connections among members of the network and identify who plays the central role.

In the context of this evaluation, SNA will build on the data collected in the country-level stakeholder survey and will provide valuable insights into the degree and nature of alignment of country-level stakeholders for the purpose of the transformative reform – one of the principles of the GPE 2025 operating model. Figure 8 provides a visual example of an SNA conducted in an evaluation context, which helps to visualize the quantified strength of relationships. The visualization will be accompanied by a narrative explanation and exploration of the relationships analyzed and, as the evaluation proceeds, we will also explore the observed changes to the network over time.
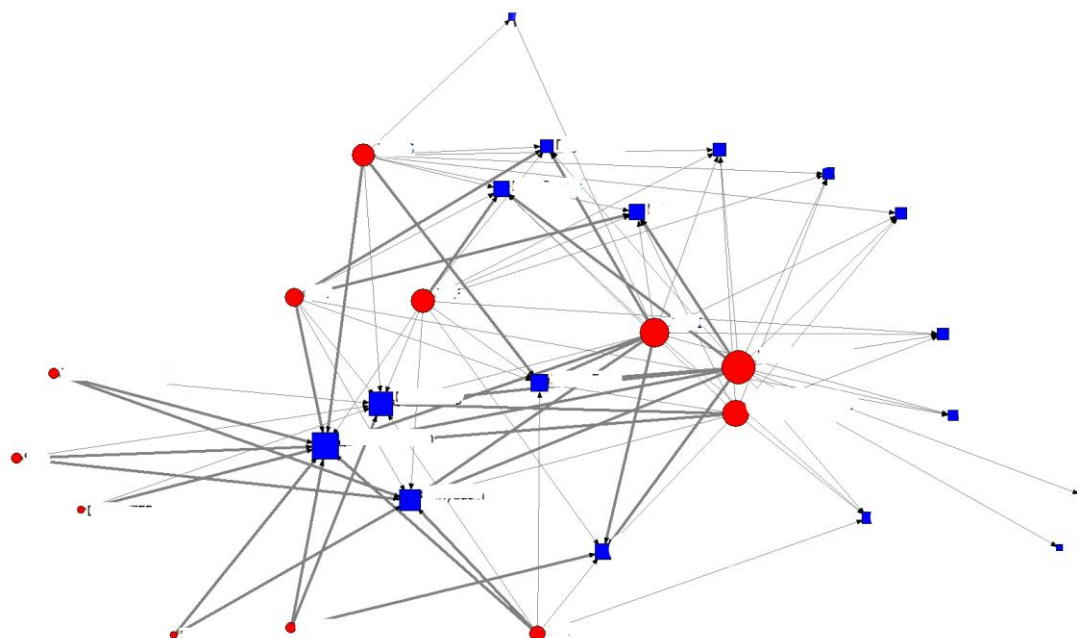

Figure 8. An example of social network analysis (visual representation of results) in an evaluation context. The network depicts working relationships between partners within a particular locale, where red nodes depict the partners and blue nodes depict locales. The size of the nodes depicts either the number of locales in which a partner is operating (red) or the number of partners in a locale (blue), while the thickness of the lines depicts the "time spent" on a defined working relationship. Note that the names of the nodes have been removed for privacy purposes.

The basis of the data for our SNA is from our country-level stakeholder survey, as the survey will serve to identify and quantify aspects of respondents' perceptions of relationships, which we will use to analyze networks at the country level rather than the characteristics of the individual stakeholders. Therefore, the unit of analysis is a country-level network. For each country-level case study, we will explore the following questions through SNA:

- How strong are the relationships among country-level stakeholders (such as those within the local education group)?
- What are the clusters of stakeholders with strong, or less strong, mutual relationships?
- How effective are these relationships in the process of compact development and its operationalization?
- Who plays a central role and who is on the periphery of the relationships?
- Who are the stakeholders through whom other stakeholders can be reached?

We will repeat the SNA in each phase using updated survey data, allowing us to track changes to relationships over time.

Gender Analysis

To support our gender hardwiring strategy, our evaluation provides analysis on gender equality on two levels: process and outcomes.

At the **process level**, we will gather and systematize evidence on how gender is hardwired in the operating model at each stage of the compact development. We will look at how GPE, including through the work of its Gender Hub, empowers and supports stakeholders in understanding the importance of gender equality to, within and through education, fostering inclusive policy dialogue, collecting disaggregated data and performing intersectionality analysis, analyzing enabling factors, identifying gender-transformative reforms and implementing gender monitoring plans to design and track gender indicators. We will also look at how GPE inputs can promote action toward gender equality outcomes in partner countries.

To assess gender hardwiring within the operating model, we will draw on elements from the Gender Continuum Framework (Figure 9) to understand the extent to which the identification, design and implementation process facilitated by GPE has been gender blind, aware or transformative.
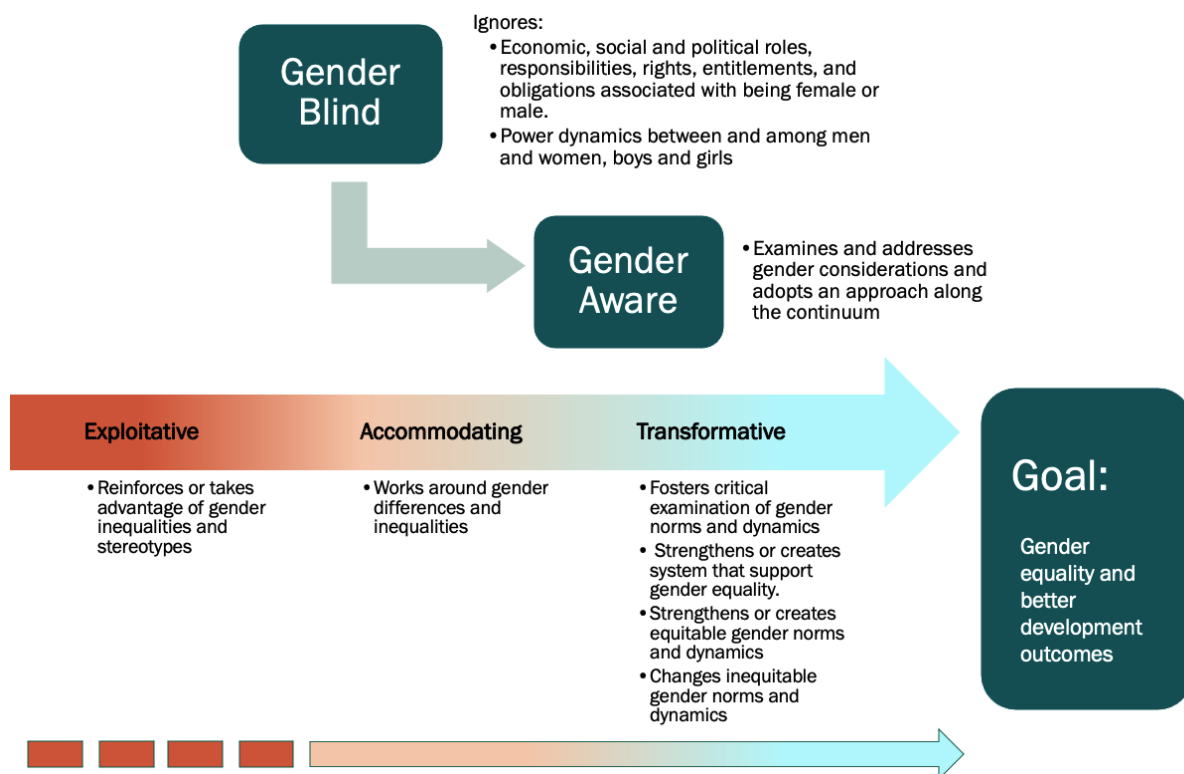
Figure 9. Gender Continuum Framework. Source: Interagency Gender Working Group

As the phases progress, we will introduce, alongside process evaluation, specific framework and questions for analysis at the **outcome level**, in which we will assess what has been achieved through the compacts in each country and assess the extent to which priority reforms have produced relevant progress in the field of gender equality to, within and through education. We will perform this analysis even if the objectives of the reform do not explicitly target gender equality or if the proposed reform is not gender-aware, as it is likely to still have consequences on access to quality learning for different groups of learners, including girls.

We will move beyond the notion of access (enrollment, transition and completion rates disaggregated by gender) to identify any progress in gender equality to, within and through education in terms of teaching and learning quality, learning outcomes, school safety and girls' and boys' empowerment (their ability to become agents of change) within society with regards to gender roles, norms and stereotypes (gender equality *through* education) and assess the contribution of GPE support to these outcomes. The assessment will rely on documentation provided by partner countries and based on how they framed and described their desired outcomes in the different components of gender equality versus what has been achieved.

As part of each country-level case study, we will perform a gender equality and social inclusion analysis on data generated at the country-level. This analysis encompasses the principle that for significant sustainable change to happen, institutional and systemic transformations related to gender and power need to occur. A transformative approach to gender and social inclusion to, within and through education focuses not only on *consequences* of gender inequality (differences in access and learning outcomes) but also on overarching structures, not limited to education, that reinforce exclusion.

We will also draw elements from the Gender and Social Inclusion Continuum (figure 10). This will allow us to assess whether the chosen reforms produced outcomes at the level of girls' practical needs or were able to address strategic needs of girls and boys, ultimately culminating in change where the broader gender relationships and power structures are addressed.
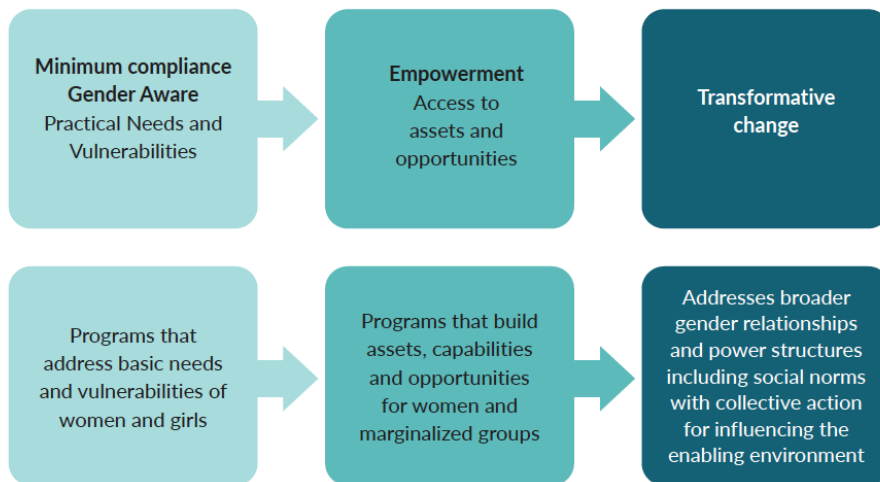
Figure 10. Gender and Social Inclusion Continuum Framework developed by Caroline Moser for the FCDO-funded Infrastructure and Cities for Economic Development Facility[59]

For each area of GPE support, we developed specific gender-focused research questions to guide our evaluation (Table 5). As with other aspects of the evaluation, these questions will be reviewed and, if required, revised between each phase of the evaluation.

---

[59] Source: CIWA/WB 2020

Table 5. Gender-focused research questions for the evaluation

Area	Process-level questions	Outcomes-level questions
Understanding the importance of gender equality to, within and through education	Is there a common understanding among stakeholders of gender equality and why it is paramount to ensure that the education systems leave no one behind? Is GPE supporting the reach of this shared understanding and how?	Has the understanding of gender equality changed as a result of the operational model processes?
Fostering inclusive policy dialogue	Are local education groups and stakeholders' groups representative of the interests of girls, marginalized groups and other minorities? Is GPE enabling processes to ensure their voices are heard? Is GPE promoting policy dialogue on gender equality to, within and through education among relevant country stakeholders?	Has the inclusiveness of policy dialogue improved as a result of GPE support? Does the implementation of the priority reform take into account the voices and interests of girls, marginalized groups and other minorities?
Data analysis, intersectionality analysis, enabling factors	Is GPE supporting local education groups, Grant Agents and countries to collect and use disaggregated data from different relevant sectors to identify and implement transformative reforms based on evidence? Is GPE supporting the capacity of Grant Agents and stakeholders to undertake gender equitable programming? Are stakeholders using GPE grants and processes to improve their gender responsiveness?	Are the priority reforms clearly addressing gender equality gaps and bottlenecks identified through data analysis, intersectionality analysis and enabling factors self-assessment?
Identification and implementation of gender-transformative reforms	Is the GPE operating model and strategy designed and resourced to tackle intersectional causes of gender inequalities to, within and through education? Can local education groups and education stakeholders identify and tackle intersectional causes of gender inequalities to, within and through education through the implementation of the priority reform? Is GPE able to influence their gender policies? Is GPE supporting in the identification of these intersectional factors? How?	How successfully were intersectional drivers of exclusion and other contextual factors that affect gender equality addressed in implementation of the reform? Who is involved in implementation? What specific outcomes toward gender equality and social inclusion were planned and to what extent have they been achieved? Which unexpected positive or negative outcomes were reached and to what extent is this attributable to the existence or absence of a prior gender equality and social inclusion needs analysis?

		Is the reform addressing intersectional drivers of exclusion beyond the education sector, requiring other actors to implement reforms in other sectors?
		Is the implementation of the priority reform resulting in progress toward gender equality to, within and through education?
		What is the potential for the implemented reform to have a transformative impact on access to quality learning for girls and boys (equality to education)?
		What is the potential for the implemented reform to have a transformative impact on gender dynamics and power structures within the education sector (equality within education)?
		What is the potential for the implemented reform to have a transformative impact on girls' and boys' capacity to be agents of change in their environment (equality through education)?
Gender monitoring and evaluation	Is GPE reviewing compacts with a gender lens? Does GPE have a reliable set of indicators to assess gender-transformative impact of selected reforms?	What indicators were measured?
	Has GPE allocated sufficient budget and resources for monitoring and evaluating the compacts with a gender lens?	
	Is GPE supporting the development of a gender monitoring system for the advancement of the proposed reform?	
	Are partner countries including a gender monitoring and evaluation system in their reforms?	

Most Significant Change

Most Significant Change (MSC) is a technique allowing for **synthesizing of impact of interventions** in very complex environments where impact and "change" may mean different things for different actors and where it is difficult to arrive at a consensus on the most important effects observed.

Developed by Rick Davies in the 1990s,[60] MSC is an iterative and participatory technique involving stakeholders and providing opportunities for arriving at a set of the most significant stories of change for which there is consensus among the stakeholders. In the context of the evaluation, MSC will allow us to better navigate through the complexities of country-level stakeholder landscapes. It is important to add that MSC will only be applied to complement – and not replace – the existing standard evaluation approaches and techniques that we propose.

We will primarily use MSC as part of the midline and endline for each country-level case study, as at baseline it will be too soon to explore and observe significant changes in partner countries arising from the operationalization of the GPE 2025 operating model. In Phases 2 and 3, we will run MSC throughout the duration of each country-level case study, from the design phase to the final analysis and reporting phase. The use of MSC will be integrated into the existing data collection tools and into the approach for the final analysis. Specific questions will be asked in key stakeholder interviews at the country level which surface and explore different perspectives on outcomes and impacts. By comparing and contrasting these subjective responses, we are able to not only triangulate any evidence of outcomes, but also further explore questions related to the dynamics of the relationships across stakeholder groups and their alignment. A separate section in each country-level case study report will report on the findings of MSC.

Triangulation

Country-level findings will be summarized against the applicable evaluation questions and subquestions, where we assess the strength of the evidence by examining the degree of agreement across data sources. Where consistencies are present across multiple sources, this helps to control for biased responses and add validity of the formulated findings. Where findings can be triangulated in our case studies, this will also deepen and improve our understanding of evaluated phenomena.

5.5.2 Cross-Case Analysis

As informed by the evaluation matrix, we will look across the findings generated from our country-level case studies as well as across our thematic case studies to conduct further cross-case analysis in support of our evaluation questions and to inform our synthesis report.

This will include using **comparative case study analysis**,[61] an analytical technique that allows us to go beyond singular, in-depth country-level examples and also examine the similarities and differences across cases. This allows us to go beyond a simple summary of findings across our cases as part of our synthesis report in order to draw out more generalizable knowledge on what works, how, why and in what ways, which supports our realist evaluation approach.

In order to utilize comparative case study analysis, following the analysis conducted at the level of the country-level case study, we will systematically review, code and extract the findings from each report. We will organize the data using a pre-determined list of variables for comparative analysis and inductively developing typologies if required. This may include comparisons across context (types of country contexts, including across partner countries affected by fragility and conflict), mechanisms (typologies for the ways in which partner countries developed their partnership compacts or have engaged with the operating model, typologies for how country-level actors have engaged or worked together, typologies of how partner countries are addressing gender equality barriers or types of priority

---

60 Davies, R. and Dart, J. (2005) *The "Most Significant Change" (MSC) Technique: A guide to Its use.* Available at: https://www.mande.co.uk/wp-content/uploads/2005/MSCGuide.pdf
61 Goodrick, D. "Comparative Case Studies." *UNICEF-Innocenti Impact Evaluation Methodological Briefs no. 9*, 2014. Available at: https://www.unicef-irc.org/KM/IE/impact_9.php

reforms) and, as the evaluation progresses, outcomes (types of outcomes, typologies for the extent to which intended outcomes were reached or types of unintended outcomes).

Following comparisons, we can then examine the nature of the similarities and differences between cases. This might include asking questions such as the following: what are the key patterns that occur over time? What might be responsible for these patterns? What is surprising about these patterns? How can similar or different patterns be explained?

We will also use comparative analysis across our country-level case studies to explore findings against GPE's key thematic priorities. As outlined in section 3.2, this will focus on gender equality (at baseline, midline and endline) and domestic financing (at baseline and endline) and as the evaluation progresses, a summary of findings against key thematic areas as they are targeted in priority reforms.[62] The analysis will help us build a picture of how the themes are conceptualized, approached and advanced at the country level. We will identify similarities and differences among the countries and, once again, identify emerging trends and patterns regarding the themes.

To support the process of cross-case analysis, we will hold internal **comparative case study workshops**.[63] These workshops will follow the analysis stages of the country-level case studies and will be conducted ahead of the cross-case analysis stage for the synthesis report. In these workshops, country-level case study teams will present their emerging findings to other country-level case study teams and to the core evaluation team. These workshops serve to support the core evaluation team members in the synthesis process by providing an opportunity to hear about and discuss the country-level emerging findings comparatively. The discussion around the findings will also explore the extent to which specific findings are based on a consensus (across cases).

The workshops will also serve an additional QA purpose for the country-level case studies as core evaluation team members are able to examine and scrutinize the data generated and findings from country-level case studies before the evaluation reports are completed. This allows opportunity for country-level case study teams to seek clarifications or follow-ups if needed and promotes consistency in the quality of country-level case studies.

Where relevant and appropriate, we will provide illustrations of the results of the comparative analysis by country-specific examples. These will be either mentioned in the text or will be inserted as concisely written (and duly referenced) illustrative text boxes into the main body of text. We may also present findings in the form of tables or matrices in a technical annex, so that comparisons are evident.

Triangulation

As with the country-level case study level, the robustness of the findings and the limitations will be duly assessed and clearly presented, in order to ensure full transparency and support an analysis of the strength of evidence toward each evaluation question, which will be presented in the synthesis reports. The findings against the evaluation questions are the starting point for identifying the most relevant issues and formulating related conclusions and recommendations.

5.6 Reporting

This section outlines the reporting process and all outputs associated with our key deliverables: **country-level case study reports** (for each country-level case study for each Phase) and **synthesis reports** (for each phase), where both cross-case analysis and the thematic analysis from across country-level case studies will be reported.

In addition to drafting reports, the reporting process will include a series of workshops and presentations which will provide opportunities to validate emerging findings – and to co-create recommendations at synthesis stage – as well as to support QA and consistency across report development. Table 6 outlines the key outputs that will be produced. Further information on the timing for each output is found in section 0. We provide some information here about the QA processes that

---

will be undertaken for each, including the role of the ITRP members James Habyarimana and Beth King. Our internal QA processes are expanded upon in section 8.5 and Annex 9.

Table 6. Key outputs produced as part of reporting for each phase

Output	Description and purpose	Quality assurance processes
Slide deck for country-level inception workshop	Slide deck that summarizes the findings of the country-level initial desk review (including PEA) and the adapted design of each country-level case study, including the country-level theory of change and adapted country-level evaluation matrix and tools. Key briefing document for country level teams.	Internal feedback to country-level case study teams from core evaluation team. Tools can be shared with R&P and country stakeholders (but not for feedback).
Slide deck for country-level debrief sessions	Slide deck that summarizes the key findings from data collection for presentation at the country-level debrief sessions. This slide deck will serve as external validation of early or emerging findings by in-country stakeholders.	Internal feedback to country-level case study teams from core evaluation team. Slide decks will be shared with R&P (but not for feedback).
Country case study reports	Report that brings together the findings and analysis for each country-level case study, including annexed PEA. Main output of the country-level case studies, on which the synthesis report will be based, to provide insights for country-level stakeholders. We will also provide 2–3 page summaries of each report for dissemination purposes.	Internal review by core evaluation team. External review by R&P (and selected GPE Secretariat stakeholders), ITRP and selected in-country stakeholders.
Slide deck for emerging findings workshop	Slide deck for presenting emerging findings from the synthesis exercise, for validation in the findings workshop ahead of report write up. To enable validation of emerging findings.	External feedback to core evaluation team from R&P and ITRP (and selected GPE Secretariat stakeholders).
Slide deck for recommendations co-creation workshop	Slide deck of proposed recommendations for discussion with R&P team as well as select number of Secretariat program specialists/managers or external stakeholders, determined with R&P. To enable inclusion in the synthesis report of agreed, practical, feasible, actionable recommendations.	External feedback to core evaluation team from R&P and ITRP (and selected GPE Secretariat stakeholders).
Synthesis report	Report that provides a cross-case synthesis of country-level case studies against key evaluation questions, presents findings against thematic evaluation questions, and recommendations. Associated synthesis report summary.	Internal review by core evaluation team. External review by R&P (and selected Secretariat or external stakeholders) and ITRP.

Output	Description and purpose	Quality assurance processes
	To provide succinct cross-country analysis against all phase evaluation questions, and clear recommendations for Board and PILC.	
Slide deck for synthesis findings and recommendation presentation	Slide deck for the presentation of complete synthesis findings and recommendations to the Secretariat Management and CEP teams, PILC and the Board. This will be built on the slide decks developed for the emerging findings and recommendations co-creation workshops. To present information clearly to enable fruitful, focused discussion.	External review by R&P (and selected Secretariat or external stakeholders).

We will produce brief **country-level reports** to summarize the findings of each country-level case study. We will produce eight baseline reports in Phase 1 (covering Wave 1, the first eight sampled countries), 15 reports in Phase 2 (midline reports for the Wave 1 and baseline reports for the seven Wave 2 country-level case studies) and 15 reports again in Phase 3 (covering Wave 1 endline and Wave 2 midline).

Reports will be concise and present findings against the relevant evaluation questions. Additional information, such as the findings from the PEA and SNA, will be included as an annex. The country-level case study reports will utilize a standard template and outline to ensure the consistency of the contents and to support comparability across country-level case studies. The QA measures will also be applied to ensure the country-level case study reports apply the overall evaluation methodology and present findings consistently.

To support the reporting process, we will conduct several types of workshops, which will allow us to test and refine our findings:

- To validate emerging findings for country-level case studies, we will first conduct **country-level debrief workshops** following data collection to support the validation of initial findings (with no analysis presented yet) with interviewees and allow for clarifications.

- Following initial analysis, we will conduct **comparative case study workshops**. These will be for the evaluation team (including country-level case study teams and the core evaluation team) to first explore the country-level case study emerging findings and to introduce a comparative lens on the findings to support the synthesis process.

- To validate emerging findings for the synthesis reports, we will hold an **emerging findings workshop** with includes our cross-case analysis across country-level case studies, and across our thematic case studies.

- In order to draft recommendations as part of our synthesis report, we will hold a **recommendations co-creation workshop** with the R&P team (with their discretion to include additional attendees). Ahead of this workshop, we will share the slide deck for the presentation of the initial conclusions for the synthesis report and emerging recommendations. In the workshop itself, the evaluation team will present the initial conclusions and facilitate a discussion in order to co-construct detailed recommendations.

We will develop a **synthesis report** for each of the three phases. The synthesis report will bring together the findings from across the country-level case study reports, presenting the key findings in an accessible manner. The structure of each synthesis report will remain similar, but the reports at each phase will explore different aspects of the GPE 2025 operating model roll-out and impact and will be guided by the phase-specific evaluation questions. As outlined in section 3.2, synthesis reports will also contain distinct chapters or annexes that focus on selecting GPE key priority themes. The contents of each report are described in Table 7.

Table 7. Synthesis report descriptions by phase

Phase	Description
Phase 1	Synthesis of findings from Phase 1 country-level case study reports against Phase 1 evaluation questions focusing on an examination of the potential for system transformation.
	Includes summary of findings across the themes of gender equality and domestic financing as either a separate chapter or annex.
Phase 2	Synthesis of findings from Phase 2 country-level case study reports against Phase 2 evaluation questions, focusing on the evaluation of the early implementation of the transformative priority reform, and also synthesizing baseline studies for the additional countries studied in the phase.
	Includes summary of findings across the theme of gender equality as either a separate chapter or annex.
Phase 3	Synthesis of findings from across the full evaluation. This will include findings from across Wave 1 country-level case study reports against Phase 3 evaluation questions (focusing on the evaluation of progress toward/achievements of system transformation and associated impact) as well as midline findings for Wave 2 countries.
	The final synthesis report will also summarize the viability of the full evaluation theory of change's results chain and set of assumptions at the conclusion of the evaluation. This will sum up key findings from across all three phases and holistically examine the strengths of linkages in the theory of change and put forward explanatory factors.
	The final synthesis report will also include a summary of findings across the themes of gender equality and domestic financing, from a summative standpoint, assessing progress throughout the operating model's roll-out and report on the other GPE priority areas, with a degree of flexibility to dive deeper into any emergent innovative themes that may not have been explicitly conceived under the original conceptualization.

All reports (country-level case study and synthesis reports) will include an executive summary, introduction, literature review, methodology and limitations, analysis, findings (and assessment of strength of evidence) organized by evaluation question, conclusions, views on further analysis required and, in the case of the synthesis report, recommendations (organized by level of priority, among others).

We will use **professional services for copy-editing, graphics production and proofreading** for all published reports. These will be provided by Scriptoria, an award-winning team of international development communications and knowledge sharing specialists that supports major development agencies and donors. Deliverables will be **translated into French**, using Scriptoria's professional translation service and the GPE Secretariat's French/English glossary of terms. To ensure that the final translations really do read as if drafted in French, a native French speaker on our team will undertake a final review to ensure that nuance is captured correctly.

5.7 Methodology-Related Limitations

In this section, we outline several anticipated practical limitations to our methodology, with a particular focus on data collection. Where relevant, we have also outlined the mitigation strategies we have developed to help reduce the impact of the limitations.

5.7.1 On Primary Data Collection

- **Progress in compact development at Phase 1:** it is critical that ahead of data collection in Phase 1, that the partner country will have completed its compact development process. We have accounted for progress against compact development as a key selection criterion for partner countries sampled for our country-level case studies. However, the data used for this selection criteria are anticipated compact development completion dates provided by GPE. In some cases,

the data on completed compact development dates may not be updated to reflect unanticipated further delays. For instance, Cambodia was selected for participation in Phase 1's country-level case studies in anticipation of the compact development process wrapping up in February 2023. However, the compact development process is now due to complete in April 2023 – while this potential delay was accounted for in the selection process and the compact will be in place ahead of the start of data collection on current plans, any further delays will directly impact the start of primary data collection. We will stagger the roll-out of primary data collection across countries to also create a buffer against any delays.

- **Research fatigue at country level:** to mitigate against research fatigue among country-level stakeholders, our selection for our sample considered countries which are participating in other recent GPE evaluations or which have heavily participated in GPE evaluations or studies in the recent past. Regardless, we recognize that there are many demands on country stakeholders' time and that the onus is on the evaluation team to ensure efficient and timely communications and minimize the demands we make. This approach will maximize trust and build relationships between respondents and the evaluation team valuable for both Phase 1 and subsequently.

- **Low stakeholder survey response rates:** consultation with the R&P team revealed that previous efforts to administer online surveys with country-level stakeholders (such as local education group members) have yielded low response rates. Therefore, we intend to conduct the survey face-to-face and will develop country-specific survey protocols at the country-level during the start-up and design phase of each country-level case study to tailor our approach to survey administration. In consultation sessions, such as with the CTL or Coordinating Agency, we may choose to further inquire about the success rates of survey administration with local education group members in the past, as there would be methodological value in expanding the survey sample. This may involve identifying the most effective modality for survey administration (e.g., online, group or face-to-face) and identifying key actors who can help promote survey engagement (such as CTLs, ministry of education representatives or Coordinating Agency staff) and developing a streamlined process for communications about the survey.

5.7.2   On Secondary Data Availability

- **Consistency in available data at the country-level:** recognizing that the compact development process, and associated efforts to compile data, can vary between countries and that countries' own capacities for collecting or collating data may differ, we may find variation in the amount of country-level data available. We will aim to use consultation sessions early in the country-level case study process and the use of templates and a unified coding framework to identify data gaps for our PEA and country-level case studies as early as possible. The early identification of gaps will mean that we will be able to consult with in-country stakeholders for solutions to data gaps over the course of fieldwork.

6   Learning Approach and Products

6.1 Learning Approach

Ongoing interaction plays a central role in our evaluation's learning approach. We aim to generate evidence and learning and engage key stakeholders (such as the GPE Secretariat, GPE Board and country-level stakeholders) in iterative feedback at the end of each phase, to inform improvements to the operating model in real time and to support GPE to achieve its strategic goals. This formative approach to learning is consistent with the new operating model's system transformation paradigm,[64] which posits that complex systems can only be understood and improved through continuous cycles of diagnosis and adaptation.

While some of our evaluation outputs, such as summaries of evaluation findings, will be planned, we will also design others based on the evaluation primary audience's emerging evidence needs. Thus, our learning approach will also adapt over the course of the evaluation.

To inform the design of our learning approach, in the inception phase, we consulted with the GPE Secretariat's 2LT and will continue to work closely with it to identify learning opportunities. We will also work with 2LT to differentiate our learning outputs from other research which may be perceived as similar by GPE stakeholders, such as ITAP's findings, ensuring that they are complementary.

6.1.1 Audiences

During the consultation sessions we held as part of our inception phase, a common theme to emerge across all stakeholder groups was of learning needs.[65] We identified the learning needs, opportunities and challenges for each of our primary evaluation audiences, summarized in table 8.

Table 8. Learning needs, opportunities and challenges by evaluation audience

Audience	Learning needs	Learning opportunities and challenges
GPE Board (including constituencies such as partner country representatives)	Priority audience for accountability and learning purposes. Requires overall insight on the roll-out of the new operating model, its relevance, efficiency and likely effectiveness. Requires evidence for strategic decision-making presented with clear trade-offs for current and future strategy.	Learning outputs and moments should be planned ahead of time with key decision-making points and regular meetings of the Board and PILC. Board and PILC meetings are short, so there may not be time for participants to reflect deeply on evidence. Evidence is presented separately for each evaluation, meaning there are fewer opportunities to assess progress holistically. There is periodic rotation among Board and PILC members, which may require socializing new ones about the evaluation study. Some Board members, including partner country representatives, may be less familiar with the operating model.
PILC	Priority audience for accountability and learning purposes.	Learning outputs and moments should be planned ahead of time with key

---

[64] However, our evaluation will maintain its independence from the GPE operating model and learning processes, for instance by providing evaluation learnings on how learning occurs within the operating model.
[65] See our report on the consultation sessions in **Annex 1** for further details.

Audience	Learning needs	Learning opportunities and challenges
	Requires overall insight on the roll-out of the new operating model, its relevance, efficiency, and likely effectiveness, with a view to identifying important areas for strategic learning by the Board. Requires inputs for making decisions as to what is to be actioned from the evaluation findings/outputs.	decision-making points and regular meetings of the Board and PILC. PILC will need additional time to reflect on findings, given its role in advising the Board on key decisions.
GPE Secretariat	Learning serves to improve the implementation of GPE's operating model. Requires insights on effectively defining and operationalizing the concept of system transformation; on how to adapt the operating model roll-out in real time; and on how to assess progress of the priority reforms, identify conditions for the successful implementation of reforms and adapt MEL to changing country-level circumstances. Requires comparative perspective on partner countries' progress against themes, including the extent to which GPE support meets country needs and is effective in each thematic area.	Learning avenues could include periodic feedback to 2LT and CEP cohort meetings, or topic-focused workshops or webinars on promising practices and country-level experiences, etc. The Secretariat comprises multiple teams, which may have varying levels of interest opportunities for engagement in different elements of the evaluation. The Secretariat is still adapting to the GPE 2025 operating model, which means that the socialization of some concepts may not yet be consistent across all teams.
Country-level stakeholders (including local education group members)	Partner countries that have completed the development of their partnership compact will benefit from learning to understand if the conditions for reform implementation are present, to maintain internal and external partner alignment throughout implementation, and to monitor and learn from implementation. Requires insights on how to engage with different features of the operating model, based on other countries' successes. Requires reflections on their own progress based on recent country-level case study visits, outputs and strategies to maximize the likelihood of effective reform implementation. Requires information on lessons and promising practices in specific thematic areas under priority reforms.	Learning moments for this purpose should be planned flexibly to account for emerging needs. Some existing venues for such learning are meetings with local education groups at the end of in-country visits, GPE partner country constituency meetings and through KIX regional hub events. Country-level stakeholders may have a lower-level baseline understanding of the new operating model. Country-level stakeholders may find it hard to convene. They likely require engagement throughout the year. Local education groups may mistakenly interpret country-level case studies as accountability exercises, as opposed to learning opportunities.[66]

---

[66] Some country-level stakeholders may benefit from learnings on country-level reform; however, this risks coming across as heavy-handed. Emphasizing learnings on GPE's support may be more appropriate though possibly less beneficial to improvements in the short term.

Audience	Learning needs	Learning opportunities and challenges
External audiences	This includes multilateral and bilateral agencies, researchers and practitioners, who may benefit from the dissemination of evaluation findings for their own research and development. This may include thematic areas or research on education system transformation. Requires insights from the operating model roll-out applicable to their own work, particularly on affecting transformative change in education systems and/or insights on progress against key priority areas.	Learnings should be disseminated, leveraging existing dissemination channels. Interests are not necessarily linked to GPE. Includes varied, niche and potentially hard-to-reach audiences.

6.2 Learning Products

As part of this evaluation, we propose to produce two types of learning products: **pre-defined, proactive** products (such as synthesis reports and presentations) as well as flexible, **on-demand reactive** products, defined in collaboration with R&P.[67] As evaluation has formative elements and intends to inform change and adaptation based on findings, we will seek to take advantage of key points in time when decisions are being made or actions are being developed to provide timely and relevant findings.

Pre-defined products, such as presentations to the Management Team of the GPE Secretariat , PILC and the Board, respond to pre-identified needs. Pre-defined learning products have been planned at various junctures of the evaluation. On-demand products are defined based on discussion of findings with R&P to assess what content is suitable for sharing with whom (both within the GPE Secretariat and externally). For example, 2LT may propose that the evaluation team produce content for the Partnership Team's newsletter.

For the **GPE Board and PILC** we will deliver the following:

- **Synthesis reports' executive summaries**: the evaluation team will draft summaries (5–10 pages) of each phase's synthesis report. These will be shared with the Board and PILC prior to relevant meetings to prepare them for discussion. Summaries will highlight findings on the operating model's implementation and will emphasize practical recommendations and potential decision-making criteria.

- **Findings presentations:** the evaluation team will produce short presentations to be presented during Board and PILC meetings at the end of each evaluation phase. These will summarize the latest evaluation findings to date and, where possible, will echo the structure of the synthesis report summaries. An outline and a draft will be shared with R&P and 2LT in the weeks leading up to these meetings to ensure alignment around key messaging.[68] These presentations will frame findings in terms of decision-making needs on the operating model both at the country and Secretariat levels.

For **in-country stakeholders**, we have identified the following products:

- **End-of-visit debriefs**: all countries will be visited twice by country-level case study teams, across phases.[69] At the end of each country visit, the country-level case study team will meet with relevant

---

[67] The current budget covers up to nine events or workshops per phase.
[68] This will account for both time required for R&P and 2LT review and alignment and time for distribution to the Board and PILC prior (two weeks in advance) to the meetings.
[69] Visits for the first eight country-level case studies will happen in Phases 1 and 3, while for the second seven country-level case studies visits will happen in Phases 2 and 3.

country-level stakeholders to share preliminary findings.[70] The workshops serve to present our main fieldwork findings back to our respondents to support their validation and allow opportunity for clarification or follow-up. The main audience of these debriefs will be stakeholders interviewed during fieldwork, but additional participants may be identified as part of the stakeholder mapping process and in consultation with CTLs or Coordinating Agencies (and other stakeholders involved in initial country-level consultations). It may also be an opportunity to share early emerging findings about operating model roll-out (in earlier phases) or achievements (Phase 3). These are intended to be active dialogues and opportunities for exploration, and not just presentations. Country-level case study teams will prepare a summary slide deck to present findings during the debrief. At the end of each phase, we will enrich the slide decks with the analysis findings and share them with the relevant country-level stakeholders. The slide decks will serve as a source of reflection and continuous learning as well as a form of progress reports to allow for cross-phase comparison.

Some learning products are likely but require further discussion. These ad-hoc opportunities will be assessed against evaluation team resourcing, availability and level of effort required, where resourcing for the pursuit of different learning opportunities will use a draw-down day approach.

- **Ad-hoc outputs:** we foresee that throughout the evaluation, R&P and 2LT may request outputs summarizing specific findings or evidence from the evaluation to be shared internally. Toward the end of the evaluation, ad-hoc outputs may be produced for a specific thematic area, such as gender.

- **Ad-hoc meetings/workshops:** short online events may also be produced for regional meetings, groups of country partners focusing on the same themes and other stakeholder groupings. These may summarize findings from relevant country-level case studies, as well as overall synthesis findings, and will mostly aim to provide comparative views on operating model adaptations and best practices among countries with similar characteristics.

Other ideas for on-demand learning products have been discussed, but are contingent on ongoing discussions within R&P:

- **Board meetings tend to have very tight agendas**, therefore **additional learning moments** outside formal Board meetings are likely necessary to **enhance Board-level learning**. For instance, **partner country meetings** prior to each Board meeting could be leveraged to ensure country representatives arrive at the Board meeting with the latest thinking on how the operating model is being implemented. A **pre-Board meeting learning day** could also be organized with the entire Board to explore the decision-making implications of synthesis report findings.

- A **separate workshop with PILC** and R&P at the end of each phase could be arranged to explore the synthesis report in depth from a methodological standpoint and to reflect on findings. This will support participants to have greater ownership of evaluation results and drive the learning conversation without the evaluators.

- **Dissemination outputs** may be produced if this requires simply adapting the style of existing learning products and adopting flexible approach to accommodate dissemination needs. Styling approach might vary based on the main subject focus, level of technicality, approach to the content delivery. Dissemination outputs will be agreed upon through discussion with GPE in different stages in accordance with the stakeholders' needs and available evaluation team resources. For instance, we will endeavor to adapt the style of our outputs to needs expressed by KIX, the Gender Hub and other GPE partners.

Learning products will thus evolve through discussion of GPE needs with both R&P and 2LT. They will be clearly defined and scheduled, and resources allocated from the ear-marked learning budget.

---

[70] These may vary across different country-level case studies, but typically include members of the local education group. The actual audience will be defined as part of Phase 1's country-level stakeholder mapping process.

7 Evaluation Workplan

7.1 Work Plan

We present a detailed, indicative workplan for Phase 1 of the evaluation (Figure 11) as well as high-level workplans for Phases 2 and 3 (Figure 12). The Phase 2 and 3 workplans will be worked out in greater detail as part of the Phase Concept Note, to be shared prior to the start of each of Phases 2 and 3.

We note that the timelines and workplan for Phase 1 will require regular updating. This is due in part to the phased approach to the evaluation roll-out in which approaches and methods are trialed and piloted and therefore subsequent timelines for activities may require revision. This is also the result of the flexible nature of the learning approach, in which there is scope to further develop and amend the timelines for learning activities in order to take advantage of ad-hoc and emerging learning opportunities with key evaluation stakeholders.

The Phase 1 workplan below will be further discussed and adapted for each country-level case study as part of the design phase, to ensure that country-level case study teams are aware of dependencies on and can indicate early on whether there may be any anticipated delays as a result of country-specific contextual factors.

Phase 1

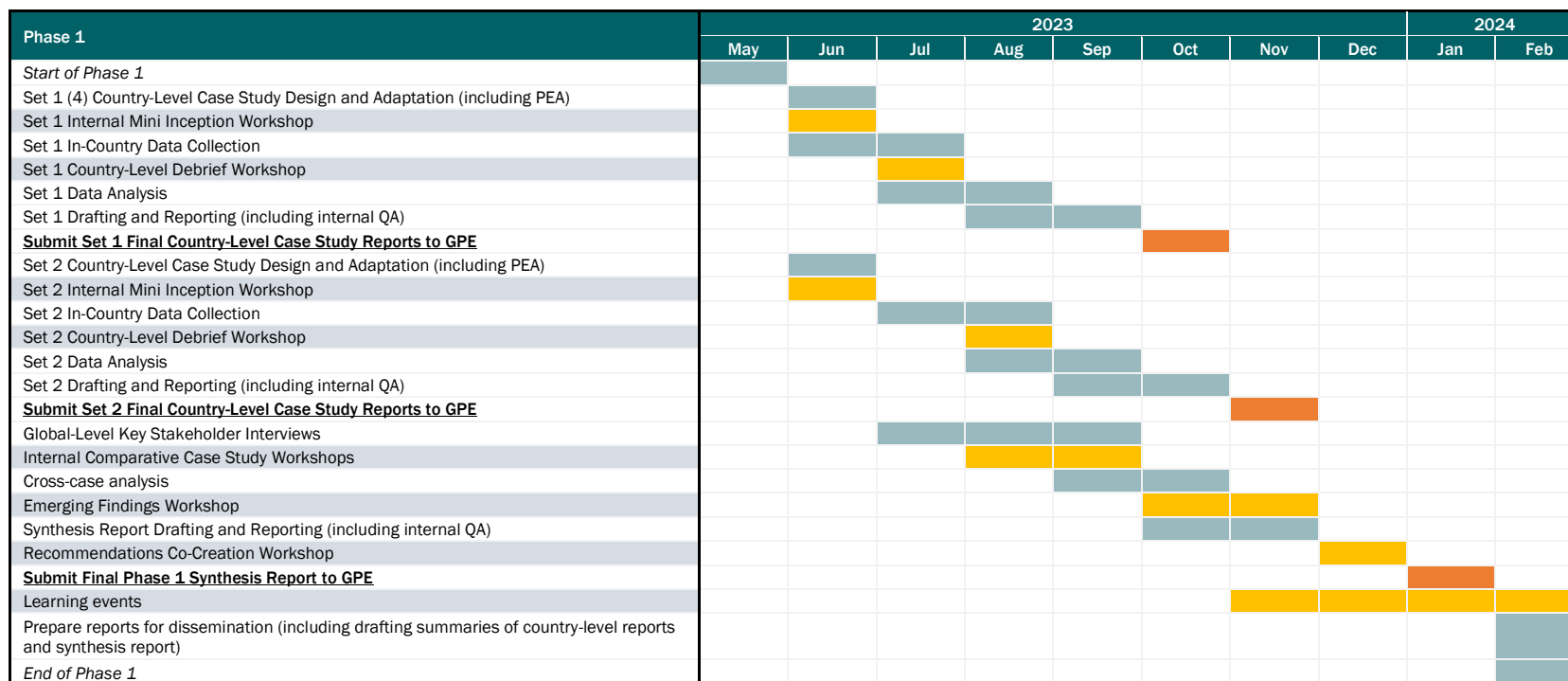| Phase 1	2023								2024	
	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Start of Phase 1	teal									
Set 1 (4) Country-Level Case Study Design and Adaptation (including PEA)		teal								
Set 1 Internal Mini Inception Workshop		orange								
Set 1 In-Country Data Collection		teal	teal							
Set 1 Country-Level Debrief Workshop			orange							
Set 1 Data Analysis			teal	teal						
Set 1 Drafting and Reporting (including internal QA)				teal	teal					
Submit Set 1 Final Country-Level Case Study Reports to GPE						red				
Set 2 Country-Level Case Study Design and Adaptation (including PEA)		teal								
Set 2 Internal Mini Inception Workshop		orange								
Set 2 In-Country Data Collection			teal	teal						
Set 2 Country-Level Debrief Workshop				orange						
Set 2 Data Analysis				teal	teal					
Set 2 Drafting and Reporting (including internal QA)					teal	teal				
Submit Set 2 Final Country-Level Case Study Reports to GPE							red			
Global-Level Key Stakeholder Interviews			teal	teal	teal					
Internal Comparative Case Study Workshops				orange	orange					
Cross-case analysis					teal	teal				
Emerging Findings Workshop						orange	orange			
Synthesis Report Drafting and Reporting (including internal QA)						teal	teal			
Recommendations Co-Creation Workshop								orange		
Submit Final Phase 1 Synthesis Report to GPE									red	
Learning events						orange	orange	orange	orange	orange
Prepare reports for dissemination (including drafting summaries of country-level reports and synthesis report)										teal
End of Phase 1										teal

Figure 11. Phase 1 indicative workplan

Note: both internal and external workshops have been highlighted and are indicated in the timeline as orange. Report submissions are highlighted in red.

We disaggregated the Phase 1 workplan by "sets" of country-level case studies, to reflect the staggered roll-out of Wave 1 country-level case study baselines. As discussed previously, each set will consist of four country-level case studies. The average completion time for each country-level case study is approximately five to six months, including finalizing the country-level case study report. Although learning events are still to be determined, we have included rows to indicate the October 2023 PILC Meetings for reference.
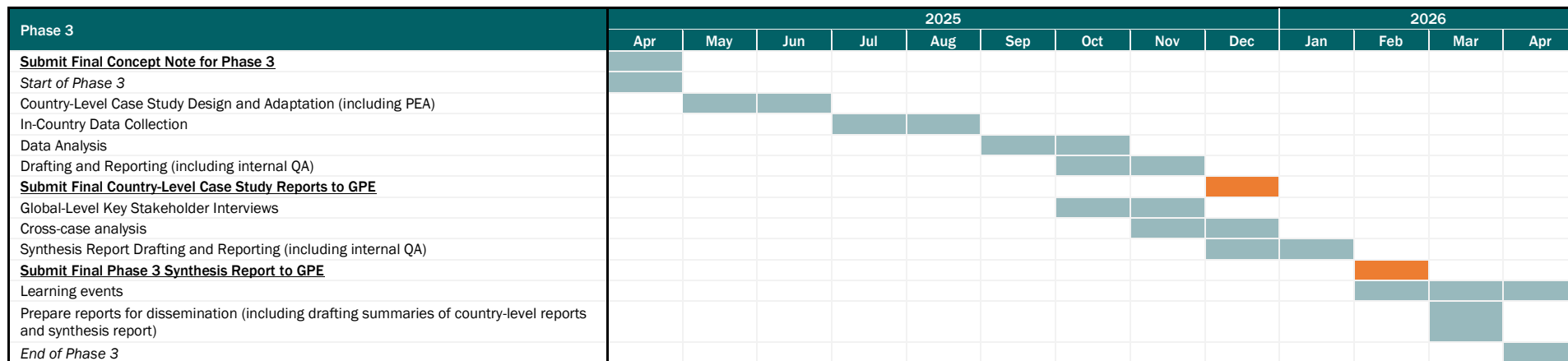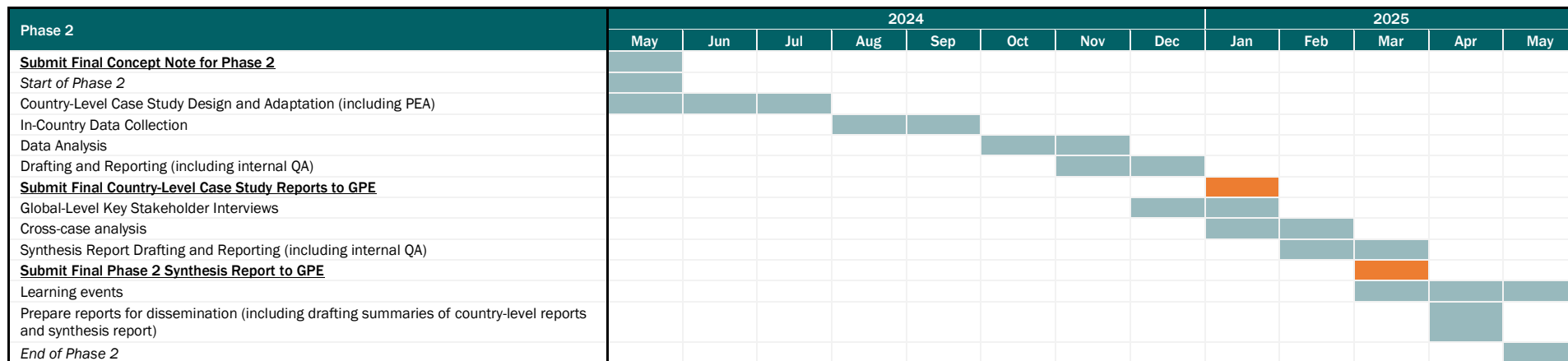
Figure 12. High-level workplans for Phases 2 and 3

Phase 2	2024								2025				
	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Submit Final Concept Note for Phase 2	X												
Start of Phase 2	X												
Country-Level Case Study Design and Adaptation (including PEA)	X	X	X										
In-Country Data Collection				X	X								
Data Analysis						X	X						
Drafting and Reporting (including internal QA)							X	X					
Submit Final Country-Level Case Study Reports to GPE									X (orange)				
Global-Level Key Stakeholder Interviews							X	X					
Cross-case analysis									X	X			
Synthesis Report Drafting and Reporting (including internal QA)										X	X		
Submit Final Phase 2 Synthesis Report to GPE											X (orange)		
Learning events											X	X	X
Prepare reports for dissemination (including drafting summaries of country-level reports and synthesis report)												X	
End of Phase 2													X

Phase 3	2025									2026			
	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Submit Final Concept Note for Phase 3	X												
Start of Phase 3	X												
Country-Level Case Study Design and Adaptation (including PEA)		X	X										
In-Country Data Collection				X	X								
Data Analysis						X	X						
Drafting and Reporting (including internal QA)							X	X					
Submit Final Country-Level Case Study Reports to GPE									X (orange)				
Global-Level Key Stakeholder Interviews							X	X					
Cross-case analysis							X	X					
Synthesis Report Drafting and Reporting (including internal QA)								X	X				
Submit Final Phase 3 Synthesis Report to GPE											X (orange)		
Learning events											X	X	X
Prepare reports for dissemination (including drafting summaries of country-level reports and synthesis report)												X	
End of Phase 3													X

Figure 12. High-level workplans for Phases 2 and 3

For the Phase 2 and 3 workplans, country-level case studies have been aggregated – these include the activities for both Wave 1 and 2 studies and present broad timelines to account for any further staggered roll-out of studies. We will discuss refining these timelines with the R&P team, including considering different timings or ways of staggering the studies to support either learning needs as they emerge or to ease the flow of studies. As with Phase 1, we anticipate that the actual completion time for each country-level case study will be approximately six months, including finalizing the country-level case study report.

7.2 Deliverables

The tables below set out our expectations for deliverables: the schedule overview (**Error! Reference s ource not found.**) and the product-specific schedule per phase (**Error! Reference source not found.** and 11). Dates for deliverables indicate the planned submission of final versions. Products and dates for Phases 2 and 3 are indicative and will be confirmed at the start of each phase.

Table 9. Overview schedule of evaluation deliverables

	Phase 1 Apr 2023 to Jan 2024	Phase 2 Jun 2024 to Apr 2025	Phase 3 May 2025 to Apr 2026
Inception report/Concept notes	1	1	1
Country-level case studies	8	15	15
Annual synthesis report	1	1	1
Dissemination and learning events (including country-level debriefs and planned presentations to GPE Management Team, Secretariat Staff, PILC and Board)	12	11	19

Table 10. Evaluation deliverables for Phase 1

Phase	Deliverables	Date
Phase 1 May 2023 to Feb 2024	**Country-level research tools** for Wave 1 country-level data collection	July 2023
	Slide decks for country-level debriefs for feedback of initial findings at the end of primary data collection in each of the eight Wave 1 country-level case studies	August 2023
	Country-level case study reports: eight brief country-level case study reports for Wave 1 countries	November 2023
	Slide deck and workshop on emerging findings at the synthesis level to be used during findings validation workshop with R&P and selected additional guests	October 2023
	Slide deck and workshop on recommendations to be used during recommendations co-creation workshop with R&P and selected additional guests	December 2023
	Phase 1 synthesis report: Synthesis of findings from country-level case studies, with chapters on thematic case studies	January 2024
	Final slide deck and presentations for Phase 1 synthesis report on findings and recommendations for Phase 1 for PILC, the Board and Management Team	January 2024
	Dissemination versions of Phase 1 reports (including 5–10 page synthesis summary and 2–3 page country-level case study report summaries for each of the eight Wave 1 studies)	February 2024

Table 11. Evaluation deliverables for Phases 2 and 3 (tentative)

Phase	Deliverables	Date
Phase 2 May 2024 to May 2025	**Concept Note** for Phase 2, setting out evaluation matrix and methodological considerations and adjustments	May 2024
	Country-level research tools for Wave 1 (eight midline studies) and Wave 2 (seven baseline studies) country-level data collection	July 2024
	Slide decks for country-level debriefs for feedback of initial findings at the end of primary data for Wave 2 (seven baseline studies)	September 2024
	Country case study reports: 15 brief case studies from eight selected baseline post-compact development countries and seven mid-term post-compact countries	January 2025
	Slide deck and workshop on emerging findings at the synthesis level to be used during findings validation workshop with R&P and selected additional guests	February 2025
	Slide deck and workshop on recommendations to be used during recommendations co-creation workshop with R&P and selected additional guests	March 2025
	Synthesis report: synthesis of findings from country-level case studies, with chapters on thematic case studies	March 2025
	Final slide deck and presentations for Phase 2 synthesis report: on findings and recommendations for Phase 2 for PILC, the Board and Management Team	March 2025
	Dissemination versions of Phase 2 reports: including 5–10 page synthesis summary and 2–3 page country-level case study report summaries for all studies conducted in Phase 2 (eight Wave 1 and seven Wave 2 studies)	April 2025
Phase 3 Apr 2025 to Apr 2026	**Concept note** for Phase 3, setting out evaluation matrix and methodological considerations and adjustments	April 2025
	Country-level research tools for Wave 1 (eight endline studies) and Wave 2 (seven midline studies) country-level data collection	June 2025
	Slide decks for country-level debriefs for feedback of initial findings at the end of primary data collection for all Wave 1 (eight endline studies) and Wave 2 (seven midline studies)	August 2025
	Country-level case study reports: 15 brief case studies from seven Wave 2 midline and eight Wave 3 endline country-level case studies	December 2025
	Slide deck and workshop on emerging findings at the synthesis level to be used during findings validation workshop with R&P and selected additional guests	January 2026
	Slide deck and workshop on recommendations to be used during recommendations co-creation workshop with R&P and selected additional guests	February 2026
	Synthesis report: synthesis of findings from country-level case studies, with chapters on thematic case studies and summative findings from the previous phases and examining holistically the strengths of linkages in the theory of change	February 2026
	Final slide deck and presentations for Phase 1 synthesis report: on findings and recommendations for Phase 3 and summative results for PILC, the Board and Management Team	February 2026
	Dissemination versions of Phase 3 reports: including 5–10 page synthesis summary and 2–3 page country-level case study report summaries for all studies conducted in Phase 3 (eight Wave 1 and seven Wave 2 studies)	March 2026

8 Evaluation Management

8.1 Governance and Reporting

The evaluation is managed by the consortium lead, Triple Line, in collaboration with Technopolis and Learn More, under the guidance of the Evaluation Manager (Jessica Chu) and overseen by the Project Director (Clarissa Poulson). The Project Director holds the main responsibility for the successful delivery of the evaluation and all contractual requirements.

Leadership and direction of the evaluation will be provided by the evaluation's **core management team**, which will consist of Michelle Kaffenberger as the Senior Technical Lead, Rebecca Allinson as the Senior Evaluation Lead and Lorenzo Newman as the Senior Thematic and Learning Lead, in addition to the Evaluation Manager and Project Director.

As well as the core management team, our evaluation team is made up of our **core evaluation team**, comprising evaluators and specialists from all three organizations within our consortium, and **country-level case study teams**. For more information on team structure and specific roles and responsibilities, see section 8.5. All members of the evaluation team will adhere to Triple Line's corporate policies and processes, and the Evaluation Manager is responsible for ensuring corporate compliance.

The consortium, with Triple Line as the main interlocutor, reports to Nidhi Khattri and Anne Guison-Dowdy from the GPE Secretariat's R&P team as primary points of contact. The evaluation team will also engage with key external evaluation stakeholders such as GPE's ITRP advisors for the evaluation, Elizabeth King and James Habyarimana, as requested and facilitated by R&P.

It is important to note that the evaluation team maintains independence from the GPE Secretariat, where the role of the Secretariat is to provide access to GPE documents and data, facilitate communications with stakeholders, provide input to guide the design and adaptation of the evaluation to ensure relevance for GPE needs, quality, appropriateness and rigor, and to ensure adherence to the design and fact-checking.

8.2 Communications Protocols

Having robust protocols and processes in place to ensure clear and streamlined communications with the GPE Secretariat and other GPE and in-country stakeholders is paramount for the smooth management of the evaluation and the delivery of high-quality products. This section sets out an overview of our protocols for meetings, communication channels and ways of working with the GPE Secretariat and external stakeholders throughout the evaluation. This is further summarized in **Annex 5**.

Email will be the main channel for communication. The Evaluation Manager, who is responsible for all key operational and management aspects, is the key point of contact for all day-to-day management and communications across our team and with the GPE Secretariat. The core management team should be included in all communication on aspects of the technical delivery of the evaluation, while the Evaluation Manager and Project Director are the key points of contact for all contractual and financial matters.

The processes that will support our approach to communications with key evaluation stakeholders are set out below.

8.2.1 GPE Secretariat R&P Team

We will hold regular meetings with members of the R&P team to communicate progress and discuss operational and technical issues. For the duration of the implementation of the evaluation, we suggest holding two types of meetings with the R&P team:

- **Weekly catch-up meetings** between the Evaluation Manager and key points of contact within the GPE Secretariat R&P team to provide regular progress updates (as well as other members from R&P and the core evaluation team, as required). These meetings will have as a standing agenda item discussions of progress against key evaluation activities and a review of timelines for upcoming activities and deliverables.

- **Monthly meetings** with the evaluation's core management team to discuss operational and technical issues and decision-making points in structured meetings with the R&P team. These meetings will be formal with an agreed and pre-distributed agenda and chaired by the Evaluation Manager, who will be responsible for note taking and sharing meeting notes following the meeting.
- Any issues that have repercussions to budget or compliance will be raised (or responded to) by the Project Director or Evaluation Manager with the R&P team directly.

GPE Secretariat or Board-level stakeholders

The evaluation team will need to communicate with stakeholders from within GPE, beyond our key contacts on the R&P team. For instance, this might include GPE Secretariat staff on the CEP team, such as CTLs. Communications will be limited to members of the core evaluation team.

- Where communication with Secretariat staff beyond R&P is required, any introductions should be first facilitated by R&P. Where contact has already been established (for instance during inception) but renewed contact is required, the Evaluation Manager and Anne Guison-Dowdy will be included in email copy in the first instance.
- Any contact with GPE stakeholders beyond Secretariat staff (for instance, with GPE Board members) should be facilitated by R&P only or, where direct contact with the evaluation team is made, R&P will be made aware of the communications and included in the discussion where relevant.

8.2.2 In-country Stakeholders

We recognize the importance of minimizing research fatigue and preserving GPE relationships and networks at the country-level, and thus the need to ensure that communications with country-level stakeholders are streamlined and initiated through the proper channels. For country-level case studies, one member of the country-level case study team will be designated as the key point-person for contact. This person will be from staff within the evaluation consortium, in the case where Country-Level Case Study Leads may be external associates. This ensures traceability and maintains accountability within the consortium.

- We will work through R&P to engage CTLs as the first step in the protocol for country-level case studies. It is expected that the CTL will then support country-level case study team engagement with Grant Agents and/or Coordinating Agencies and other local education group members as relevant.
- We anticipate that Grant Agents and/or Coordinating Agencies are likely to be conduits to further in-country stakeholders, which may include line ministries beyond the education sector (such as ministries of women and children and ministries of finance), civil society, education stakeholders (such as teacher organizations) and policy makers (including those who work outside education ministries). More information on how stakeholders will be engaged throughout data collection (e.g., as participants) is outlined in sections 5.4 and 5.6.

8.3 Risk Management

Effective management of risk underpins all aspects of our evaluation. Robust risk management will ensure that work is delivered on time in a way that ensures the highest standards of probity and safety for all involved, wherever it is undertaken. We will apply Triple Line's established **risk management approach** (see **Annex 6**) which we have tailored to this evaluation.

We have also outlined the anticipated risks for this assignment in our **risk matrix** (see **Annex 7**), which includes risks related to the operating environment (contextual risk), outputs (delivery risk), how country-level case studies are delivered (safeguarding and operational risk) and how the evaluation is managed (fiduciary and reputational risk). The risk matrix outlines all the key risks identified by the evaluation team, providing a description of each type, its likelihood of occurrence and subsequent impact. Our risk matrix also includes a short description of the mitigation plan the team has or will aim to employ, should such risks arise. Mitigation actions are assigned to a team member to ensure ownership.

All members of the evaluation team have and will contribute to risk identification and analysis, applying their different areas of expertise and experience as well as contextual knowledge to ensure that all

types and aspects of risk are considered. Further contextual risk assessment will take place at the start of each country-level case study. This will include a security risk assessment, which we will undertake together with Triple Line's network of security companies such as International SOS.

The risk matrix will be reviewed bi-weekly by the Evaluation Manager and the select members of the core evaluation team. Some risks will be tolerated; others will be treated (their likelihood or impact reduced), terminated (avoided through a change in approach, e.g., related to geography) or escalated for further discussion. The results of this review will be disclosed to the R&P team during the joint weekly check-ins. Where the core management team deems relevant, escalating risks and suggested responses will be raised with R&P.

Where new or escalating risks are identified (e.g., risks that are increasing in likelihood and/or impact), this will be brought to the attention of the core management team either through email or in regular meetings, and communicated to country-level case study teams as appropriate.

8.4 Ethical Conduct, Safeguarding and Data Protection

The evaluation's research ethics, safeguarding and data protection policies are rooted in our responsibilities to treat all those with whom we work with dignity and respect. We recognize that we act from a position of privilege and that unequal power dynamics limit the voice and agency of many. We will utilize a number of tools and policies to ensure that our team is abiding by the highest standards of ethical and safety conduct. We summarize some of the key principles here and provide greater detail on each topic in our research protocols (see Annex 8). Our protocols for this evaluation are based on Triple Line's policies and procedures on research ethics, Code of Conduct, Child Protection Policy, Safeguarding Policy, Data Security Protection and Cyber Security Policy, and Guidance on Reporting Concerns (whistleblowing). Copies of each can be found attached to our research protocols. These policies apply to all Triple Line staff members and all those subcontracted by Triple Line, either directly (e.g., consortium members and associates) or indirectly (e.g., associates of consortium members).

8.4.1 Ethical Conduct and Safeguarding

Our evaluation team recognizes the importance of ethical standards and safeguarding in research and aims to maintain the highest professional standards in these areas. To support this, our research protocols present a common set of principles and procedures, which guide the actions of all those engaged in our evaluation team in supporting ethical conduct and the prevention of harm. Our research protocols also present our policies and procedures to prevent sexual exploitation, abuse and harassment, which are designed to protect adults and children (men, women, girls and boys) from any harm that may be caused in their interactions with Triple Line or anyone engaged with our evaluation.

As part of the on-boarding process for country-level case study teams, we will provide training on our research protocols and require all teams to be familiar with and agree to adhere to all Triple Line policies. Associated sexual exploitation, abuse and harassment and safeguarding risks have also been identified in our risk matrix and will be regularly monitored as part of our risk management approach.

8.4.2 Duty of Care

As the consortium lead, Triple Line holds the responsibility for duty of care provisions to support the evaluation team, and notably, for in-country visits. The work and travel of the evaluation team is governed by Triple Line's Duty of Care Policy. The key components of our duty of care include the conduct of country-specific risk assessments prior to any country visit, conducted by Triple Line's operations team with support from our retained security consultant and based on the latest information from sources including the UK's FCDO.

This will inform the drafting of risk management plans tailored to the risk environment in each country to be visited. These will include security planning to cover general and specific risks associated with staff travel, work environment, accommodation and training requirements. We will also hold pre-deployment meetings to brief those involved in country-level case study data collection on the risk context and their role in supporting the ongoing monitoring of risks. Live and ongoing risk monitoring will be supported by our external security partner, and in-country teams will be required to report on a regular (e.g., daily) basis. In cases where there are considerable risks for in-country data collection, such as if sites cannot be accessed due to high security concerns for researchers, visits will not be

conducted and contingency plans for remote data collection will be developed. In these cases, workplans may need to be re-sequenced to avoid unmanageable risks.

8.4.3 Data Protection and Management

Our research protocols also include information on how we will protect and manage data. All members of the evaluation team will abide by Triple Line's overarching Data Protection & Cyber Security Policy. This policy details how we will manage and protect all data shared by GPE and collected during the course of this evaluation. Information will also be made available to those from whom we collect data as part of our data collection consent forms (also included in our research protocols).

Our research protocols also set out our data management principles and processes for how we will ensure that all data are handled with the utmost care and responsibility in order to avoid loss, damage or inappropriate access, including compliance with applicable legal, regulatory and international obligations. We also include information on how data will be de-identified, secured and disposed.

8.5 Quality Assurance

We have robust standards and processes in place to ensure the high quality of our deliverables. We have prepared a full QA framework (see Annex 9) to set out our internal processes and protocols for supporting the quality of methods and tools (including evaluation design fidelity), quality of data and evidence, and quality of outputs and deliverables throughout our evaluation. This allows us to ensure that this evaluation will meet GPE's standards (as outlined in GPE's Evaluation Policy) and fulfill the terms, conditions and expectations specified in the contract, evaluation terms of reference and present inception report.

To support our QA framework, we have integrated the responsibilities for QA across the evaluation team. This allows us to delegate QA responsibilities at the right levels and in a manageable way by sharing the effort across the team, supporting the right people/team members to conduct QA checks and ensuring that QA is done in the same way in order to promote both quality and consistency. In Table 12, we provide an overview of the roles for QA, which are described further in Annex 8 9.

Table 12. Internal quality assurance roles and responsibilities

Roles	Quality assurance responsibilities
Country-level case study teams	Country-level case studies are carried out in accordance with the overall evaluation approach and with contextual insight
	Country-level case studies are carried out in a way that accounts for the local political economy and with respect to local culture and norms
Core evaluation team	Country-level case studies are completed with methodological fidelity and rigor, with consistency across studies
	Ensure all country-level case studies are carried out to ethical research and safeguarding standards
	Ensure that thematic analysis and synthesis reporting are carried out in accordance with the overall evaluation approach and are duly reflective of the findings stemming from the country-level case studies
Core management team	Provide high-level feedback throughout the country-level case study process, through engagement at the country-level mini inception workshop, in-country debriefs (virtually) and internal synthesis workshop
	Review of all deliverables
Triple Line in-house support	Proofreading and editing for informal deliverables (those that will not be published)
Scriptoria	Proofreading and editing for all formal deliverables that will be published

The Evaluation Manager is responsible for ensuring that the QA processes are socialized across the core evaluation team and all country-level case study teams and monitoring and supporting compliance. The Evaluation Manager is also responsible for ensuring that QA process is regularly updated, informed by discussions with the R&P and the wider evaluation team as the assignment progresses. Learnings from early QAs about how country-level case studies have been undertaken will be fed into subsequent studies and this feedback loop will further ensure quality.

8.6 Team Structure and Management

Given the extensive coverage and longitudinal nature of the evaluation, it is important that we have robust processes in place to guide the management of our evaluation team within and across the three phases, up to 2026. This allows us to ensure that the right expertise is utilized at the right times, and we are able to ensure smooth continuity and learning throughout the evaluation.

To support this, we use a **team of teams approach** to staffing, where we designate clearly defined teams with responsibilities for different aspects of the evaluation. This flexible team of teams approach avoids the pitfalls of a more traditional and cumbersome hierarchical structure, which typically involves silos and ends up leaning too heavily on individual team members, creating risks in cases of team replacements. Instead, responsibilities are shared, and collaboration is encouraged across a group of individuals who are able to contribute their specialized knowledge and experience.

- A centralized **core management team** is responsible for leading our work, bringing together the expertise of our evaluation leads with key management functions and sharing the responsibility for designing the evaluation and ensuring that the evaluation is implemented with fidelity.

- Furthermore, our **core evaluation team** brings together methodologists, thematic experts and managers to carry out the main functions of the evaluation. This core team brings coherence and consistency to the evaluation by bringing together involvement in the areas of country-level case studies, thematic case studies, synthesis and learning. Our core evaluation team also includes our Gender Lead, who in addition to leading the design and conduct of our evaluation activities related to gender equality, will support the team in ensuring that gender is hardwired throughout all evaluation activities. Core evaluation team members will be internal staff members of our consortium, ensuring greater consistency and reducing the likelihood of turnover. Core evaluation team members will also fulfill more than one functional role in the evaluation (e.g., also taking on roles in the country-level case studies and working across thematic case studies and synthesis and learning activities), thus increasing consistency in approach, improving knowledge exchange and reducing risk.

- Finally, small and clearly defined **country-level case study teams** will deliver each individual evaluation, thus enabling lean, agile delivery and continuous learning. These teams include a balance of members with country-specific contextual knowledge and experience and core evaluation team members rooted in the evaluation design. The involvement of our core evaluation team members (as Qualitative Evaluators as well as Research Analysts) helps to ensure both consistency across country-level case studies as well as continuity for the team.

The membership of the teams is described in Table 13 and our team organogram and specific roles and responsibilities are found in **Annex 10**.

Table 13. Description of the evaluation teams

Team	Purpose	Members
Core management team	Management and oversight Technical design and direction of the evaluation Quality assurance	Senior Evaluation Lead (Rebecca Allinson) Senior Technical Lead (Michelle Kaffenberger) Senior Thematic and Learning Lead (Lorenzo Newman) Project Director (Clarissa Poulson) Evaluation Manager (Jessica Chu)

Core evaluation team	Day-to-day technical delivery, including oversight of country-level case studies and delivery of thematic analysis, synthesis reporting and learning activities	Evaluation Manager (Jessica Chu)
		Country-Level Case Study Lead (Adam Krcal)
		Country-Level Case Study Manager (Ekaterina Shaleva)
		Gender Lead (Alice Pelosi)
		Domestic Finance Lead (TBD)
		Learning Manager (Giovanni Zino)
		Quantitative Evaluation Lead (Giorgio Monti)
		Data Manager (James Handley)
		Qualitative Evaluators
		Research Analysts
Country-level case study teams	Delivery of country-level case studies	Case Study Leads
		Qualitative Evaluators
		In-Country Researchers
		Research Analysts

Our team of teams approach, in which teams are well integrated and overlap, allows us to support the continuity of our work in the event of staff turnover. Where staff turnover occurs, we will firstly (and most immediately) look to fill the role from within our existing team (such as filling a gap in a country-level case study team with someone involved in either the core evaluation team or another country-level case study team). Where this is not possible, or where a longer-term solution is required, we will fill roles from in-house staff and experts across the consortium or across trusted associates where a particular expertise is required and not available in-house.

Strong internal communications protocols will help us to establish internal cohesion between teams as well as ensuring that evaluations are both consistent in adhering to the overall frameworks and approaches, and sufficiently tailored to the subject matter. Internal communications and tools for team management include the following:

- **Microsoft Planner, dedicated SharePoint and Teams channels:** as all three consortium members make use of the Microsoft 365 Suite, we are able to use dedicated Microsoft solutions to support evaluation management. These include the following:
    - o **Dedicated SharePoint site:** we have a dedicated SharePoint site for the evaluation, which allows us to securely store all evaluation data as well as create a space for collaboratively working on documents and outputs. As set out in our research protocols, access will be granted to our SharePoint site as necessary, meaning that team members will only be granted access to folders as required.
    - o **Dedicated teams channels:** we will also set up dedicated teams channels for each of our country-level case study teams, allowing for ease of communication beyond weekly meetings and emails. This also allows the core evaluation team to disseminate information quickly and easily to each of the country-level case study teams.
    - o **Microsoft Planner:** this tool enables the core evaluation and core management teams to quickly track and monitor progress against country-level outputs. Each required output for the country-level case studies is set up as a "task," where the core evaluation team can ensure that tasks are assigned to relevant team members. Country-level case study teams will use the tool to provide short updates on progress toward the task and raise issues. The Country-Level Case Study Manager is responsible for ensuring that Planner is regularly updated by teams.
- **Weekly meetings (country-level case study teams):** for the country-level case studies, each country team will hold weekly team calls, chaired by the Case Study Lead. Country-level case study teams will be expected to discuss ongoing technical activities and planning, and reflect progress made on

evaluation outputs in Planner. These meetings will be held for the operation duration of country-level case studies.

- **Weekly meetings (core evaluation team):** the core evaluation team will meet on a weekly basis to review progress made across country-level case studies, and share issues arising and learning which can usefully be applied or cascaded out to all case study teams. They will also discuss technical activities and progress related to the thematic case studies, synthesis and learning phases. The team will review the progress as reported in Planner and identify any issues that should be escalated to the core management team, such as changes to the workplan, changes in the status of risks or changes to team resourcing and personnel. Where relevant, Case Study Leads will also participate in these calls, which will serve to provide rapid reporting upwards and provide an opportunity for cross-study learning.

- **Weekly meetings (core management team):** these meetings serve as an opportunity for the core management team to be regularly appraised of evaluation progress and discuss the issues that have been raised by the core evaluation team as they emerge. any issues raised that have repercussions on budget or compliance will then be escalated by the Project Director or Evaluation Manager to R&P.

- **Regular review of workplans and team resourcing/availability (Evaluation Manager):** regular review and refreshing of individual country-level case study workplans and the overall evaluation workplan as progress is made, to ensure expert availability for the tasks envisaged. Triple Line's resource planning platform Illuminis provides at-a-glance visibility of team commitments and availability. Ensuring availability of staff is a key aspect of our project continuity planning.

Details on the weekly meetings and internal team communications are also outlined in our communications protocols (**Annex 5**).

Annexes

Annex 1. Inception Phase Consultation Report

Introduction

The Global Partnership for Education (GPE) 2022–2026 Study: Country-Level and Thematic Evaluation is being undertaken by Triple Line and consortium partners Technopolis Group and Learn More with the following aims:

- Examining how GPE's global and country-level processes and instruments support results at the country level

- Identifying promising practices and explaining their underlying success factors in context

- Generating recommendations to strengthen GPE's operating model and strategy to support transformative reform processes.

The inception phase began in November 2022. Throughout, the evaluation team have worked closely with the GPE Secretariat Results and Performance team (R&P), who manage the evaluation on behalf of the GPE Secretariat. A critical aspect of the inception phase has been to consult with a wide range of GPE stakeholders, both within and outside of the Secretariat, in order to ensure that design the evaluation meets the needs of its primary users. R&P, as gatekeeper to these stakeholders, arranged and facilitated these sessions in order to do the following:

- Introduce the evaluation consortium and key team members

- "Socialize" the evaluation across GPE, including to raise awareness on its objectives, timelines and key outputs

- Share the approach taken by the evaluation

- Allow opportunities for stakeholders to provide feedback and ask questions to inform the evaluation.

A consultation slide deck was developed in discussion with R&P and tailored for each session held.

Formal consultation sessions were followed by two whole days of in-person workshops facilitated by the evaluation core team and attended by GPE Secretariat stakeholders and the evaluation's independent advisory team. More information on the attendees of each can be found at the end of this report.

This report gives a short account of the consultation sessions held during the inception phase and the main points which arose from them, and likewise for the workshops. These points have fed into evaluation design, which will be set out in the evaluation inception report.

Consultation Sessions

Six consultations were held between January and February 2023 with the following stakeholder groups:

- **Leadership Group of the GPE Secretariat Country Engagement & Policy (CEP)** which acts as GPE Secretariat's primary relationship coordinator with country partners and is key to implementation of the GPE's new operating model, which is the subject of the evaluation.[71]

- **GPE Secretariat management team**, responsible for steering and delivering GPE's work, comprising of the Executive Office (Chief Executive Officer (CEO), Deputy CEO, Chief of Staff) and managers from across the Secretariat teams.

- **Constituent groups of the GPE Board**, responsible for reviewing performance of key strategic initiatives that are core to GPE's goals, considering the key risks that can derail those goals and

---

[71] CEP supports national evidence-informed policy dialogue at country level, including on the development and implementation of inclusive Education Sector Plans, Partnership Compacts and Joint Sector Reviews to accelerate transformative progress in policy priorities. Source p2 of: https://www.globalpartnership.org/docs/jobs/GPE-Inclusive-Internship-Program-2022-Policy-Strategy-CEP.pdf. Retrieved April 4, 2023.

adapting and adjusting its business model and platforms as warranted.[72] This includes donors, partner countries and representatives from civil society organizations (CSOs), the private sector, private foundations and multilateral agencies.

Full details, including the session dates, target groups and all attendees, are found in Annex 2.

Consultation Findings

Although sessions were not recorded, for the purpose of encouraging open discussion, comprehensive notes were taken at each session. We also invited consultation participants to share written feedback and follow-up questions with us.

We captured the key points discussed in these sessions and in written feedback in a data matrix. Setting the points out systematically enabled us to consider each one and record our internal response to it as our design for evaluation methodology developed. The matrix approach also enabled us to analyze all the points raised by sorting them into categories, thus making visible the frequency with which they were raised, and by whom.

The following sections set out the comments received, organized by category of comment, the group of consultees who provided the feedback and how we will respond to the comment in our evaluation.

Evaluation Structure, Methodology and Timing

There were 19 points raised by stakeholders on implementation of the evaluation – its focus, scope and methodology (including sampling). The table below summarizes these comments in sub-categories, indicates their origin and gives our overall response.

Summary of points raised	Consultee	Response
Importance of considering the role of context, particularly fragility and conflict, and assessing the performance of the Operating Model in emergency settings, as well as in the COVID context	GPE Board (CSO/private sector/foundations/multilaterals; donors) GPE Secretariat (CEP leadership group)	Evaluation already includes country-specific political economy analysis to surface contextual differences and account for fragility/conflict and implications. Evaluation will use a realist evaluation approach to also consider context. Country sampling (primarily for Phase 2 of country-level case studies) to include fragile and conflict settings.
Need for country visits and importance of robust methodology (particularly where country-level data are scarce or unreliable)	GPE Board (donors; partner countries) GPE Secretariat (core management team)	Methodology (and budget) will now include country visits in order to strengthen evidence and deepen country-level case studies. Detailed information on data collection methodology and data quality assurance (QA) processes to be included in inception report.
Sampling considerations to avoid bias (e.g., due to roll-out progression) and research fatigue	GPE Board (donors)	Both are taken into account as sampling criteria. For Phase 1, sample is limited due to roll-out progress, but in future phases the sample will be re-balanced to account for model roll-out.

---

[72] https://www.globalpartnership.org/who-we-are/board. Retrieved 4 April 2023.

Summary of points raised	Consultee	Response
Scope – how will the evaluation ensure depth, including to look at both process and impact at the country level	GPE Board (CSO/private sector/foundations/multilaterals; donors)	Country-level case studies will be in-depth and include more resourcing. Our phased evaluation approach will focus on impact at the country-level in later phases of the evaluation; this will be explained in the inception report.
Importance of breadth of engagement at the country-level – including development partners, civil society, government/ministries	GPE Board (partner countries)	Country-level case studies will include stakeholder mapping and initial consultations to ensure that a broad range of voices are included in data collection.

Learning, Communications and Dissemination

Stakeholders were asked to provide feedback on how to support the usability of the evaluation, particularly with regard to learning and dissemination. We received 17 comments from stakeholders (summarized into categories below) on concerns and suggestions for how the evaluation can support engagement.

Summary of points raised	Consultee	Response
Clarity on the role of the Board and in particular PILC in reviewing and engaging with the findings	GPE Board (donors) GPE Secretariat (CEP leadership group)	Key evaluation outputs will be made publicly available and presented to the Board (including PILC). R&P team are responsible for involving and managing GPE Board's engagement with evaluation deliverables.
Timing – important that the evaluation is able to report findings and lessons learned regularly and ahead of the end of the evaluation	GPE Board (donors)	Phased approach used by evaluation will report findings approximately yearly prior to the completion of the evaluation.
Importance of ensuring the accessibility of findings – in terms of format (both technical/non-technical, long/short, distill contents/learnings by audience), language and modality (e.g., different media)	GPE Board (donors)	Evaluation team to work closely with R&P and GPE Secretariat to tailor outputs appropriately to audience, using existing GPE channels.
If findings are to be shared through online forums (e.g., webinars), there needs to be adequate time built in for discussion and questions	GPE Board (donors)	Evaluation team to work closely with R&P and GPE Secretariat to ensure that online presentations are appropriately structured.
Evaluation findings should be disseminated at the country-level including, if relevant, to provide practical guidance for engagement with GPE	GPE Board (partner countries)	Evaluation team to work with R&P and CEP to ensure that findings are appropriately disseminated at country-level,

Summary of points raised	Consultee	Response
		including the use of post-mission debriefings.
Importance for the evaluation to contribute to strengthening global evidence base, including ensuring the evaluation's relevance to global discussions (such as global events) and to contribute to GPE's wider position/reputation	GPE Board (donors)	Dissemination for contribution to GPE's global engagement is beyond the remit of the evaluation, but the evaluation team will work closely with R&P and GPE Secretariat to support the relevance and timing of any dissemination related to the evaluation.

Key Areas of Investigation for the Evaluation

Consultations yielded a number of useful comments on various facets of the GPE operating model and GPE's non-financial support. Some of these comments can be classified as "early evidence" (e.g., providing helpful indications and feedback from stakeholders on the performance of the operating model thus far) and have contributed to the evaluation team's understanding of how the model operates. Examples include anecdotes about the challenge of the operating model roll-out (particularly accounting for different contexts), the time and effort required thus far in the compact development preparations, the benefits of the "priority setting" phase and the process for enabling factors analysis.

Other comments will help direct the evaluation toward key areas of investigation. All these points have implications for the evaluation theory of change and its underlying assumptions, as well as the evaluation's lines of inquiry (including evaluation questions and subquestions, as laid out in the evaluation matrix). We have incorporated these considerations into our design of the evaluation's theory of change (including assumptions) and evaluation matrix. In the following section, we provide a short description of these comments, organized by areas of investigation for the evaluation.

On the GPE Operating Model

We received 34 comments which relate to questions on the design of the GPE 2025 operating model, its roll-out and the different actors involved. Additionally, we received a small number of important comments on theories of systems transformation, including on GPE's conceptualization and definition of it and approach to it, within design of the program.

Summary of points raised	Consultee	Response
The operating model process from the design perspective: the application process, fund disbursement, efficiency, role of ITAP; reporting and success factors for how the model operates; and whether early reflection on the model has resulted in adaptation and learning and how the evaluation findings will remain relevant given that some adaptations will have been made after the relevant processes were completed in countries in the Phase 1 sample	GPE Board (CSO/private sector/foundations/multilaterals; donors) GPE Secretariat (management team; CEP leadership group)	Built into the theory of change and assumptions, and evaluation matrix including mapping adaptations made to the operating model to evidence challenges on the operating model.
How is GPE defining transformation, what are indicators of success and what	GPE Secretariat (management team; CEP leadership group)	System transformation was a significant topic in the co-design workshop that the evaluation

Summary of points raised	Consultee	Response
was being promised by it? Noting the new approach being taken in the GPE 2025 operating model and the thinking being developed by external stakeholders		team facilitated with stakeholders within the GPE Secretariat. The work presented and built on the discussions from the consultation sessions and a detailed desk review of the literature. The inception report will describe the criteria (attributes of system transformation) that will be used in the evaluation.
The process of compact development at the country-level: the design process including country ownership/leadership, involvement by different actors within the country-level ecosystem (including on the role of local education groups and civil society and their inclusion); the availability, reliability and use of data; how the grants are used, including whether compacts have informed grant implementation or brought more financing; the implications of the COVID-19 pandemic for model roll-out; and how the model is supporting transformational change	GPE Board (CSO/private sector/foundations/multilaterals; donors; partner countries) GPE Secretariat (core management team; CEP leadership group)	Built into the theory of change and assumptions, and evaluation matrix.
In-country actors, their capacities and influence on the compact development and grant implementation – capacity constraints of government, the role of local education groups (and their inclusion) and the selection and performance of Grant Agents	GPE Board (CSO/private sector/foundations/multilaterals) GPE Secretariat (core Management team	Built into the theory of change and assumptions, and evaluation matrix. Accounted for as part of stakeholder mapping to inform sampling of respondents at the country-level. Addition of an in-country survey as part of the methodology to investigate the role of different actors at the country-level and on the effectiveness of alignment through social network analysis.

On the Added Value of GPE and the Performance of the Partnership Model

Nine comments, almost all made at the Board Constituencies meetings, were made in relation to the added value of GPE to a country's policy making (an important premise for the new operating model is that it enables governments to own and drive transformation) and to the landscape of development interventions in education.

We have separately identified four comments specifically on the performance of the partnership model which underpins GPE's operating model and strategy. These points were made in consultations held with the GPE Secretariat core management team and the GPE Board (donors and multilateral agencies).

Summary of points raised	Consultee	Response
Unintended consequences included government funding being displaced by GPE funding, or alternatively compact implementation absorbing other available funding	GPE Board (CSO/private sector/foundations/multilaterals)	Built into the theory of change and assumptions, and evaluation matrix.
Value add of the partnership – both assessing this and disseminating it more widely, beyond GPE	GPE Board (CSO/private sector/foundations/multilaterals)	Built into methodology – including social network analysis and proposed evaluation matrix; internal dissemination planned.
Coordination with other initiatives and fit in the ecosystem, including at the country-level	GPE Board (CSO/private sector/foundations/multilaterals) GPE Secretariat (core management team)	Will be an element of the political economy analysis undertaken at the start of each country-level case study.
The effectiveness of GPE's partnership model, how to define effective partnerships (including use of effectiveness partnership principles) and what it means and to whom	GPE Board (CSO/private sector/foundations/multilaterals) GPE Secretariat (core management team)	Partnership is built into the theory of change and developed as an underpinning concept for the evaluation, building on work previously undertaken on this topic by the evaluation team on other programs, which are grounded in a partnership model.
Examination on how partnerships work on the ground, how this might support transformational change and the supporting factors to effective partnership working	GPE Board (donors)	Questions about partnership working have also been included in the proposed evaluation matrix.

On Key Thematic Priority Areas, Including Gender Equality and Inclusion

Stakeholders raised five points on thematic areas of interest for the evaluation. These included interests in equity and inclusion as well as on gender quality.

Summary of points raised	Consultee	Response
Ensure that the evaluation focuses on the way in which gender is hardwired in the GPE model	GPE Board (donors) GPE Secretariat (core management team)	The evaluation's approach to evaluating gender hardwiring, as well as the evaluation's own approach to mainstreaming gender equality, will be explicitly described in the inception report.
Importance of exploring the effectiveness of the role of GPE in supporting vulnerable and hard-to-reach populations (including adolescent girls, displaced populations, out-of-school children)	GPE Board (CSO/private sector/foundations/multilaterals; donors; partner countries)	Evaluation has a revised approach to key thematic areas. This includes a focus on gender equality and social inclusion throughout the evaluation, and will explore key thematic areas as they arise in priority reforms in our country-level case studies.

Workshops

The evaluation workshops were designed to enable a "deep dive" into fundamental components of the framework for the evaluation – the theory of change (including its underpinning assumptions and concepts relating to systems thinking and transformation) and methods (evaluation questions and methodology including sampling and learning). The purpose was to discuss, add to and validate the evaluation team's conceptual thinking so that they could further refine their design of the evaluation. Attendees on both days can be found at Annex 3.

Day One – Theory of Change

Systems thinking and systems transformation was the topic of the first presentation. Points raised, including from the Independent Technical Review Panel members present, included whether to encompass all schools and not just publicly funded schools in thinking about the education system in a given country; the importance of context, actors, relationships and above all leadership; the boundaries of analysis; and the role of innovation within a system.

Theory of change and its underlying concepts followed. The theory of change conceptualized a pathway up to compact development and then a pathway using the compact as the starting point to meeting chosen objectives – ultimately, education systems transformation and SDG 4.[73] Questions raised echoed points made during the consultation sessions: the extent to which the new model represented genuine ownership or simply meeting compliance requirements; the importance of the evaluation looking at the new approach being taken by the new operating model; the difference between country-level outputs and outcomes compared to global impact; that gender needed to be prioritized throughout; and the importance of partnerships and their underlying dynamics.

Assumptions underpinning three successive elements of the theory of change were discussed in breakout groups as follows, starting with those aspects of the theory of change over which GPE has greatest control:

- Development of the compact with its focus on priority reform

- Operationalization of the compact to achieve priority reform, supported by the enabling factors and including alignment of funding

- Wider education systems transformation.

Discussions interrogated the assumptions, their rationale and means of testing them. Feedback was captured and subsequently fed into revision of the assumptions.

Thematic case studies and their usefulness for learning globally and at the country level was the topic of the final discussion of the day. Points raised included the objectives of looking at themes as well as country-level reform in the evaluation, noting that countries would have differing interests but were keen to learn from each other; and those themes varied – some could be seen as inputs (e.g., domestic financing), others as outcomes (e.g., inclusion) or as cross-cutting (e.g., gender). Discussions on Day Two confirmed that gender and domestic financing would be evaluated across the country-level case studies, for inclusion in the synthesis reports, not as standalone thematic reports.

Day Two – Methods

Evaluation questions were presented under successive elements of the evaluation: the process of compact development; post-compact development country-level case studies (Phase 1a); post-compact development thematic case studies at the country level (Phase 1b); and post-compact development thematic case studies at the global/cross-national level. Feedback was sought on specific proposals and some points on **evaluability** and related issues arising from consultation sessions (such as data availability and research fatigue) were shared and discussed.

---

[73] Sustainable Development Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.

Methods and the main principles in implementing them, with a focus on Phase 1, were presented and discussed. The **sampling approach** was outlined, sampling considerations and criteria were applied to the 12 eligible countries for Phase 1, and the sample was selected[74] in discussion for GPE approval.

Learning principles and ways of working were presented for discussion. Discussion covered the needs of different stakeholders, particularly when and how to engage country level actors, the purpose of learning and respective roles of the evaluation team and R&P (including the Learning Team).

Conclusion

Our consultations yielded a number of important areas of concern from key evaluation stakeholders. This provided the evaluation team with important refinements to the evaluation's approach, scope and methodology (including on learning and dissemination) as well as on key areas of investigation. Below, we provide a summary of the key takeaways across each of the areas identified above:

- All stakeholders emphasized the importance of ensuring that contextual, country-level analysis is included as part of the evaluation, with a particular focus on the ways in which fragility and conflict affect partner countries' engagement with the GPE operating model. We have accounted for this by using political economy analysis and detailed stakeholder mapping for the start of each country-level case study, which will influence the adaptation of research tools and the identification of respondents for data collection. We also added the use of an in-country stakeholder survey and social network analysis to investigate country-level actor dynamics. To ensure that our country-level case studies include countries in later cohorts of the model roll-out as well as fragile and conflict-affected countries, we use a phased approach to sampling, beginning with eight countries studied in Phase 1 and an additional seven countries for Phase 2. The evaluation team also agreed with R&P on the addition of in-country field visits as part of the methodology to strengthen contextual data collection. This supports additional depth for the country-level case studies as well as breadth and diversity of voices to be included. These approaches will be outlined in greater detail in the inception report.

- Consultations provided a number of considerations for the learning needs of various stakeholders, including to inform the Secretariat's work and global engagement as well as at the country-level. To support the use of findings, we will work closely with R&P and team members across the GPE Secretariat to ensure that we are disseminating findings in an accessible format/medium as well as to account for timing (feeding into decision-making and events). The inception report will outline our approach to learning and dissemination for Phase 1, which will adopt a flexible and collaborative approach as the evaluation team and R&P work together to determine the most effective means to support learning.

- Stakeholders provided a large amount of feedback on initial observations of the GPE operating model roll-out and suggested a number of areas for investigation in the evaluation. Areas included the design and implementation of the GPE operating model, as well as GPE's non-financial support (through GPE's partnership model and value add) and support to GPE's key thematic priority areas. We will ensure that these observations and feedback are embedded into our understanding of the GPE operating model and support in the design of the evaluation's theory of change (and assumptions) as well as evaluation matrix (in terms of evaluation questions and subquestions). Our inception report will provide greater detail on the design and key aspects of our theory of change and how it frames our theory-based evaluation, as well as on our evaluation matrix. The inception report will also include details on how gender equality and social inclusion are mainstreamed throughout our evaluation, both in terms of the evaluation methodology and its lines of inquiry on GPE's gender hardwiring process.

- Stakeholders were interested to know how the evaluation will account for the adaptations that have already taken place to the operating model and whether the evaluation can continue to inform adaptations and improvements to the model. The phased approach of the evaluation will allow the evaluation team to examine different "versions" of the operating model, by sampling

---

[74] Cambodia, Democratic Republic of the Congo, El Salvador, Nepal, Tajikistan, Tanzania – Mainland, Sierra Leone and Uganda.

across different cohorts of operating model for our country-level case studies. We will also examine the effectiveness of the adaptations by comparing findings across phases. Our phased approach will also allow us to generate learning feedback more quickly, reporting findings at the end of each phase (and therefore roughly on an annual basis). The phased approach and our approach to learning will both be described in the inception report.

The workshops provided a valuable forum to raise awareness about the evaluation with GPE stakeholders, build common understanding about it and provide opportunity and space for comment and discussion. They thus helped to further refine the evaluation team's approach to the conceptual underpinnings of the evaluation and its design.

Annex 1.1 Consultation sessions schedule and attendees

(All sessions additionally attended by members of the GPE Secretariat R&P team)

Date	Consultation group	Attendees	Attendees from the evaluation team
January 10, 2023	GPE Secretariat: Leadership Group of the Country Engagement and Policy Team	- Jo Bourne; GPE Chief Technical Officer - Sally Gear; Gender Hub Lead - Morgan Hanadi Strecker; CEP Strategy and Coordination Lead - Hadi Khan; Programme and Planning Officer - Raphaelle Martinez; Education Policy and Learning Team Lead - Fazle Rabbani; Regional Lead - Tahinaharinoro Razafindramary; Regional Lead - Nilse Ryman; Regional Lead	Rebecca Allinson, Jessica Chu
January 17, 2023	GPE Board: CSOs, private sector, private foundations and multilateral agencies (Part 1)	- Solange Akpo, ANCEFA - David Boutcher, Reed Smith LLP - Flavia Brunetti, WFP - Abeer Darwazeh, ACEA - Giorgia Maddalon, WFP - Dennis Sinyolo, Education International - Rebecca Telford, UNHCR	Rebecca Allinson, Jessica Chu, Mahima Mehra, Lorenzo Newman, Giovanni Zino
January 18, 2023	GPE Board: donor partners	- Ilham Alghadiri, UAE - Juanita Botha, Canada - Charlotte Busert, Germany - Annica Floren, European Commission - Camilla Fossberg, Norway - Anna-Lena Flury, Switzerland - Louise Holt, Canada - Marianna Knirsch, Germany - Charlotte Lebas, France - Lauren Luckhurst, UK - Jeff Mettille USA - Jessica Purdie, UK - Elin Ruud, Norway - Rouksana Simjee, France	Rebecca Allinson, Jessica Chu, Adam Krcal, Mahima Mehra, Lorenzo Newman, Clarissa Poulson

Date	Consultation group	Attendees	Attendees from the evaluation team
		• Dan Thakur, Canada • Wouter Van Damme, European Commission • Emily Woolf, UK	
January 25, 2023	GPE Board: CSOs, private sector, private foundations and multilateral agencies (Part 2)	• Natalia Cheratova, World Bank • Harriet Nannyonjo, World Bank	Rebecca Allinson, Jessica Chu, Lorenzo Newman, Giovanni Zino
February 9, 2023	GPE Secretariat: core management team	• Jo Bourne, Chief Technical Officer (Country Engagement and Policy Team) • Julie Desangles, Chief of Staff • Margarita Focas Licht, Manager (Partnerships Team) • Laura Frigenti, Chief Executive Officer • Raphaelle Martinez, Education Policy and Learning Team Lead (Country Engagement and Policy Team) • Charles North, Deputy Chief Executive Officer • Nidhi Khattri, Manager (Results and Performance Team) • Matthew Smith, Risk and Compliance (Finance and Grant Operations Team) • Charlie Tapp, Interim Chief Operating Officer (Secretariat Operations Team)	Jessica Chu, Clarissa Poulson
February 14, 2023	GPE Board: partner countries	• Dr Hari Lamsal, Nepal • Ms. Phumzile Magagula, Eswatini	Rebecca Allinson, Jessica Chu, Giovanni Zino

Annex 1.2 Workshop Attendees

Both sessions, held in the World Bank in Washington DC, were attended by the following:

Evaluation Team:	Rebecca Allinson	**R&P Team:**	Nidhi Khattri
	Adam Krcal		Anne Guison-Dowdy
	Jessica Chu		Rudraksh Mitra
	Lorenzo Newman		Booyoung Ko (virtual)
	Clarissa Poulson		Gauri Khanduja

Independent Technical Review Panel for the evaluation: Elizabeth King

Thursday 23rd February

Attending for all or part of the day:

> Jean-Marc Bernard, R&P and Learning Team
> Aya Kibesaki, CEP (CTL for Rest of the World)
> Sally Gear, CEP (EPL Gender Lead)
> David Balwanz, Finance and Grant Operations (Quality Assurance Senior Education Specialist)
> Raphaelle Martinez, CEP (EPL Team Lead)
> Tahinaharinoro Razafindramary, CEP (Regional Manager for francophone Africa)
> Nilse Ryman, CEP (Regional Manager for Rest of the World cluster)
> Blandine Ledoux, CEP (CTL for francophone countries)
> Morgan Hanadi Strecker, CEP (Senior Policy and Planning Officer) (virtual – Paris)
> Yuliya Makarova, Finance and Grant Operations (Monitoring Officer)
> Plamen Danchev, CEP (CTL for non-francophone countries)
> James Habyarimana (Independent Technical Review Panel for the evaluation)

Friday 24th February

The evaluation team was joined by Michelle Kaffenberger

Annex 2. Theory of Change Assumptions

Assumption	Indicators	Data sources	Analytical tools
Assumptions that cut across the theory of change			
1 The composition of the **local education group** is inclusive of different voices, including of civil society and teacher groups.	The mapping of the composition of local education groups (over time) shows inclusivity of different voices Share of consulted stakeholders who: • Consider inclusive policy dialogue an important part of underpinning principles in the GPE operating model • Feel their voices are sufficiently heard • Feel others' voices are sufficiently heard. Satisfaction level among local education group members with the inclusive nature of policy dialogue Identification of relevant stakeholders who are not represented Evidence of mechanisms allowing to seek input from relevant stakeholders who are not represented in local education group	Stakeholder mapping Document review Survey of stakeholders at the country level Interviews or focus group discussions with stakeholders at the country level, including with policy makers Expert interviews	Social Network Analysis Qualitative content analysis Triangulation of data
2 A critical mass of **country-level partners** (including Grant Agents and Coordinating Agencies) is active and remains fully **committed** to and **demonstrates**: (1) working together, under government leadership, during the compact development process and operationalization of the priority reform (including by avoiding aid fragmentation), (2) learning together/continuously adapting and (3) being held accountable (including around gender mainstreaming).	Share of consulted country-level partners who: • Believe that system transformation (that is the principles underlying system transformation such as alignment of actors) is an appropriate approach to reforming the education system • Demonstrate commitment to the compact development (e.g., by committing to actions for alignment in compacts) and express commitment to the operationalization process • Express their commitment to adapting their priority reform linked programs based on evidence from their implementation • Express their commitment to mutual accountability during the compact development and prioritization process, and throughout reform implementation • Agree the priority reform is acceptably resourced	Stakeholder mapping Document review Survey of stakeholders at the country level Interviews with stakeholders at the country level	Qualitative content analysis Social Network Analysis Political economy analysis

Assumption	Indicators	Data sources	Analytical tools
	Share of consulted stakeholders who: • Agree that the process is driven by government • Agree that all relevant stakeholders are involved and there is buy-in • Appreciate input from international donors and other international partners to the process. The mapping of the stakeholders involved in the compact development process, operationalization and wider system transformation in partner countries (national/regional/local) shows diverse stakeholder engagement and outlines their roles in the process		
3 At country-level, there are **appropriate incentives, functioning mechanisms and sufficient capacity** (or plans to address capacity gaps) (including GPE actors – Grant Agents, Coordinating Agencies etc.) at the country level to: (a) co-develop, operationalize the priority reform, (b) avoid fragmentation of aid, (c) learn together/continuously adapt and (d) being held accountable (including around gender mainstreaming). Additional required support is available to **partner countries affected by fragility and conflict** (PCFC).	Country partners have sufficient capacity and mechanisms in place to: • Fully engage in the process of compact development and operationalization (and avoid aid fragmentation), including on mainstreaming gender throughout • Learn together and adapt programs and policies continuously • Be held mutually accountable. Evidence of required support available to PCFC for the process of compact development and operationalization, learning/adapting and mutual accountability (including for mainstreaming gender)	Document review Interviews or FGDs with stakeholders at the country level, including with policy makers Expert interviews	Qualitative content analysis Political economy analysis
4 The **GPE Secretariat** remains **committed** to the GPE 2025 operating model and gender hardwiring within the model, and to continuously learning/adapting as a means of supporting wider education system transformation, and it has the **capacity** to support its roll-out and implementation.	A vast majority of the consulted GPE Secretariat staff members express their continuous commitment to the GPE operating model, to hardwiring gender to achieve results and to continuous learning/adapting Evidence of continuous sufficient human, technical and financial capacity within the GPE Secretariat (including panels and systems in place) to supporting the model roll-out, including in relation to learning/adapting and gender hardwiring	Document review Interviews with GPE Secretariat staff members	Qualitative content analysis
5 After accounting for GPE grants and funding from other donors, **domestic finance** remains adequate to deliver priority reforms.	Evidence that after accounting for external finance, domestic financing is available in sufficient quantity and is equitable	Document review, including on public expenditure	Qualitative content analysis

Assumption		Indicators	Data sources	Analytical tools
		Evidence that GPE grants and other donor funding do not lead to reduction in domestic financing levels to an extent that would impede reform implementation	Interviews with stakeholders at the country level Expert interviews	National statistical data analysis Political economy analysis
6	There are signs that beliefs and perceptions on the most important **purpose of school** (and education in general) are converging among the relevant stakeholders and align to the objectives of the priority reform.	Beliefs and perceptions about the most important purpose of school (including on gender equality in and through education) are converging in the partner country and align to the objectives of the priority reform	Document review Literature review on beliefs and perceptions on the most important purpose of school (and education in general) as well as gender equality in/through education (+ examples of usual beliefs and perceptions, including on gender) of various stakeholders (e.g., parents, teachers, policy makers, school leaders and pupils) Interviews or FGDs with stakeholders at the country level Expert interviews	Qualitative content analysis Political economy analysis
7	There are functioning mechanisms in place to address usual **resistance to reform.**	Evidence of mechanisms in place to address usual resistance to reform Evidence that countries are taking/plan to take action to address usual resistance to reform, and with what level of success Share of consulted stakeholders who: • Are aware of the risk related to resistance to reform • Agree that the partner country needs to take appropriate action (and who indicate what such action could look like).	Literature review to identify common ways in which reform is resisted (e.g., fear of loss of jobs, fear of pay cuts, perception of loss of status/power and perception of increased pressure on pupils/students, teachers, parents, school leaders, regulators etc.) Country document review to assess whether mechanisms are in place to address resistance; and if addressed, whether such resistance is eventually remediated Interviews or FGDs with stakeholders at the country level Expert interviews	Qualitative content analysis Political economy analysis

Phase 1 – Laying the foundations and preparing for the reform

Assumption		Indicators	Data sources	Analytical tools
8	The **enabling factors, ITAP processes, compact development processes** (including data disaggregation and gender-responsive sector planning, policy and monitoring) and **system capacity grant** are well **understood**, considered **relevant, sufficient** as a means of solving bottlenecks and strengthening capacity, and **enable good policy dialogue and capacity building** on system capacity gaps.	Share of consultees: With a similar understanding of enabling factors, ITAP, compact development and system capacity grant processes Agreeing that the selected enabling factors and ITAP processes are relevant and sufficient to identify critical bottlenecks to transformation in the partner country, and the system capacity grant to strengthen system-level capacity in the three dedicated areas Evidence that the enabling factors, ITAP and compact development processes enable good policy dialogue on system capacity gaps, and the system capacity gaps strengthens system-level capacity in the three dedicated areas	Interviews or FGDs with stakeholders at the country level	Qualitative content analysis
Phase 2 – Enacting the reform and Phase 3 – Reaching results through the reform				
9	The **GPE operating model and strategy and related mechanisms** mobilized to facilitate reform implementation and continue capacity strengthening (e.g., system transformation grant, girls' education accelerator, multiplier and system capacity grant) and other compact partners' support, are well **understood** and considered **relevant, sufficient and cost-effective** as a means of implementing the priority reform.	Share of consultees: With a similar understanding of GPE operating model and strategy and related mechanisms during reform implementation, e.g., System Transformation Grant (STG), Girls' Education Accelerator (GEA), Multiplier Grant (MLT), SCG, etc. Agreeing that this implementation support (GPE + compact partners) is relevant, sufficient and cost-effective to yield transformation results and strengthen capacity in the partner country	Interviews or FGDs with stakeholders at the country level	Qualitative content analysis
10	**Adequate and timely data/evidence** at the country-level is **available** and appropriately **used** to allow for monitoring and adaptation during the operationalization of the compact, including the implementation of the enabling factors and the priority reform (including gender-disaggregated data, gender analysis and intersectional factors analysis).	Existence of a results framework/compact periodic review which is regularly updated, and the results shared with all engaged actors in a timely way Share of consulted stakeholders who (a) use the evidence produced, (b) agree that decisions during the compact operationalization are made on the basis of evidence (including gender-disaggregated data, gender analysis and intersectional factors analysis) and (c) agree that the decisions/adjustments made are appropriate based on the identified cause(s) of the issue	Desk review of compact mid-term reviews Desk review of partners' monitoring reports on reform progress, including government annual reports (sector and other) and donor program reports Interviews or FGDs with stakeholders at the country level	Qualitative content analysis Political economy analysis

Assumption		Indicators	Data sources	Analytical tools
11	GPE operating model and strategy is being adapted based on **emerging emergency circumstances** (such as natural disasters, health emergencies and conflict).	Evidence of provisions in the GPE Operating Model allowing to accommodate countries affected by emergency circumstances	Document review	Qualitative content analysis
		Evidence that partner countries affected by emergency situations were promptly and suitably supported to meet unexpected needs and reassess or maintain planned reform objectives	Interviews or FGDs with stakeholders at the country level	Political economy analysis

Annex 3. Country-Level Case Study Sampling Methodology

Background

This annex describes the sampling strategy for the 15 country-level case studies undertaken as part of the Country-Level and Thematic Evaluation of the GPE 2025 Strategy and Operating Model. This overall sampling approach is applicable for all 15 cases, but this annex will only describe the process for the selection of the first eight country-level case studies, which will begin as part of Phase 1 of the evaluation. Prior to the start of Phase 2, we will revisit this sampling methodology and assess its effectiveness, as well as undertake the selection process for the next seven country-level case studies which will roll-out in Phase 2. This will be further described in the concept note prepared ahead of Phase 2, which will begin in May 2024.

The evaluation's terms of reference specify that the study will sequentially examine GPE's operating model and strategy from the beginning of the roll-out of the new GPE model (2021) up until mid-2025. The evaluation will use 15 country-level case studies as longitudinal studies through which we will assess the (continuous) relevance and coherence, efficiency, effectiveness and (potential for) impact of GPE operating model and strategy in helping partner countries achieve a chosen transformative reform, its associated impact and potential for sustainability. Country-level case studies will involve both desk-based and remote research as well as primary data collection at country level (in-country or otherwise).

The new GPE operating model is being rolled out in phases, with a total of five planned cohorts. The first "pilot" cohort was rolled out in January 2021, with the fifth cohort scheduled to begin in April 2023. Further details are included in the table below.

Cohort # (number of partner countries)	Cohort start date	Countries
Cohort 1 (Pilot) (6)	January 2021	Democratic Republic of the Congo, El Salvador, Kenya, Nepal, Tajikistan, Uganda
Cohort 2 (14)	October 2021	Bhutan, Burkina Faso, Cambodia, Comoros, Fiji, Gambia The, Senegal, Sierra Leone, Somalia – Somaliland, Somalia – Puntland, Somalia – Federal, Tanzania – Mainland, Tanzania – Zanzibar, Zimbabwe
Cohort 3 (16)	January 2022	Benin, Chad, Côte d'Ivoire, Guyana, Kiribati, Maldives, Marshall Islands, Federated States of Micronesia, Nicaragua, Niger, Rwanda, Samoa, Sao Tome and Principe, Solomon Islands, Tonga, Tuvalu, Vanuatu
Cohort 4 (10)	July 2022	Ethiopia, Cabo Verde, Djibouti, Eritrea, Guinea, Lesotho, Liberia, Madagascar, Nicaragua, South Sudan
Cohort 5 (31)	January 2023	Afghanistan, Bangladesh, Burundi, Cameroon, Central African Republic, Democratic Republic of the Congo, Dominica, Ghana, Grenada, Guinea-Bissau, Haiti, Lao People's Democratic Republic, Malawi, Mali, Mauritania, Mozambique, Myanmar, Nigeria. Pakistan - Balochistan, Pakistan – Khyber Paktoonkhwa, Pakistan – Punjab, Pakistan – Sindh, Papua New Guinea, St. Lucia, St. Vincent and the Grenadines, Sudan, Syrian Arab Republic, Timor-Leste, Togo, Republic of Yemen, Zambia

As the new GPE operating model is currently being rolled out in phases, the evaluation will adopt a phased approach as well, to be coherent with the model roll-out. The inception report explains in further detail the way in which the country-level case studies will be rolled out: it will be done in two waves, starting with eight countries in Phase 1 ("Wave 1") and an additional seven countries in Phase 2 ("Wave 2"). This staggered approach will allow us to begin our country-level case studies as countries are completing the development of their partnership compact, allowing us to examine this process in detail as part of our "baseline" study.

Overall Sampling Approach

The purpose of the sample will be to ensure that the evaluation considers a diverse set of GPE partner country contexts in which GPE's operating model (and both financial and non-financial support more broadly) is being implemented, in order to understand and generate learnings on the conditions in which the GPE operating model and strategy has or has not been successful to support partner countries.

To achieve this, we propose a **purposive and stratified sampling approach** that will be applied across a select number of sampling criteria, alongside the status of a partner country with regards to its engagement with and participation in the new GPE operating model. We draw primarily on the data collected by the GPE Secretariat to track the operating model roll-out (the "OM Pipeline"), as well as a selection of secondary data on countries' education systems and other criteria in order to examine the diversity of contexts represented by the sample.

Sampling Criteria

To define the "universe" or "population" of countries eligible for this study, we use the list of countries eligible for STGs as a proxy for the countries who will be engaged in the GPE 2025 operating model, as eligible countries are required to have developed a partnership compact. In total, 77 partner countries[75] are anticipated to be eligible for funding. It should be noted that the evaluation will also consider countries that are not eligible for STG but are eligible for the GEA grant and therefore undertake the development of a partnership compact as part of the requirements for access.[76]

We used the following inclusion criteria:

- **Geographic diversity** (using GPE constituencies)

- Diversity of countries based on status with regards to fragility and conflict, using GPE's designation as a **PCFC**

- Diversity of countries based on **income group** (using World Bank designations)

- Diversity of **country political context** (unitary vs. federal government structure; government effectiveness score using the World Bank Worldwide Governance Index)

- Diversity of **country education system context, structure and status** (including considerations for percentage of all children in early childhood/primary /secondary education and percentage of girls in primary/secondary education, using data from the GPE Results Framework)

- Diversity of **cohort designation** (with regard to the operating model roll-out)

- Diversity of **GPE funding mechanisms and maximum allocation size**.

We also accounted for the following as exclusion criteria:

- **Research fatigue:** countries that are already engaged in a number of ongoing GPE evaluation activities

- **Compact development estimated date:** countries will be required to complete their compact development process prior to the start of our evaluation phase, in order for our country-level case studies to have a process to evaluate. For the purpose of the evaluation, we assigned a **cut-off date of June 2024**, as this is when Phase 2 of our evaluation and second (and final) wave of baseline studies is scheduled to begin.

We anticipate that a total of 52 partner countries will be eligible for our sample, based on the scheduled date for their completion of the compact development process (as of February 2023); there will be 25 countries that fall outside of our window.[77]

---

[75] This is fewer countries as in the case of Tanzania, Somalia and Pakistan, regions within the country are eligible for GPE grant support. This includes: Tanzania – Mainland, Tanzania – Zanzibar, Somalia – Federal, Somalia – Somaliland, Somalia – Puntland, Pakistan – Balochistan, Pakistan – Khyber Paktoonkhwa, Pakistan – Punjab and Pakistan – Sindh. Each of these regions is included in the total figure of 77 partner *territories*, a total which is drawn from 71 different *countries*.

[76] There are 22 countries which are ineligible for STG but eligible for MLT alone or MLT and GEA.

[77] These 25 countries are Afghanistan, Burundi, Cameroon, Central African Republic, Democratic Republic of the Congo, Dominica, Grenada, Guinea-Bissau, Lao People's Democratic Republic, Malawi, Mali, Mauritania, Myanmar, Nigeria, Pakistan – Balochistan, Pakistan – Khyber Paktoonkhwa, Pakistan – Punjab, Pakistan – Sindh, St. Lucia, St. Vincent and the Grenadines, Sudan, Syrian Arab Republic, Timor-Leste, Togo and Zambia.

Wave 1 Selection Process

The selection for Wave 1 countries took place as part of the Theory of Change and Methodology Co-Design Workshop, held in-person at the GPE Secretariat offices in February 2023. Members of the evaluation's core management team and the R&P team from the GPE Secretariat were in attendance, as well as the members of the Independent Technical Review Panel (Elizabeth King and James Habyarimana).

The start of Phase 1 was scheduled for March 2023; therefore we applied a **cut-off date of February 2023 for compact completion** for eligibility for Wave 1 selection. As a result, 12 countries were eligible for the sample.[78] The full table of Wave 1 sample-eligible countries against our sampling criteria is found at the end of this annex.

The evaluation team presented the sampling criteria indicated above as well as the resulting sample frame. The group discussed the merits of each partner country's participation as a case study for this evaluation, agreeing that eight countries was a suitable number of countries to select. The group agreed on the following:

- Selecting between the two Latin American and the Caribbean countries, El Salvador was selected over Guyana on the basis of El Salvador being a unique example of a country not eligible for STG but having still elected to complete the compact development process.

- The GPE constituency "Africa 1" was over-represented across the 12 eligible countries. Tanzania – Mainland was selected over Tanzania – Zanzibar and Uganda was selected over Zimbabwe on the basis of the likely larger size of its grant allocation, and therefore the extent of GPE's financial support. Finally, Kenya was not selected due to its heavy participation in other recent evaluations.

As a result, the group all agreed on the following eight countries: **Cambodia, Democratic Republic of the Congo, El Salvador, Nepal, Tajikistan, Tanzania, Sierra Leone and Uganda.**

Next Steps

Ahead of the start of Phase 2, and the selection of seven further countries, we will revisit this sampling approach in order to ensure its suitability for identifying a diverse set of country cases through which to examine and evaluate the GPE 2025 operating model and strategy.

We will further update our understanding of the roll-out of the operating model in order to develop the full list of countries eligible for selection as part of Wave 2. We will examine the degree to which our Wave 1 countries are balanced against our stratified sampling criteria and determine against which sampling criteria our sample needs to be further balanced.

Ahead of Phase 2, we may also consider the following sampling criteria:

- Examining a diversity of Grant Agents who support partner countries in the implementation of their chosen priority reform (through the STG)

- Examining a diversity of GPE financial instruments which a partner country has accessed

- Examining a diversity of key priority areas in which the chosen priority reform of a partner country falls.

---

[78] This information was provided by the GPE Secretariat to the evaluation team.

Wave 1 sample-eligible partner countries by GPE constituency, PCFC status and income group

Country	GPE engaged since	GPE constituency	PCFC	Income group
Cambodia	2006	Asia and Pacific	No	Lower-middle income
Congo, Democratic Republic of	2012	Africa 2	Yes	Low income
El Salvador	2022	Latin America and the Caribbean	No	Lower-middle income
Nepal	2009	Asia and Pacific	No	Lower-middle income
Sierra Leone	2007	Africa 3	No	Low income
Tajikistan	2005	Eastern Europe, Middle East, Central Asia	No	Lower-middle income
Tanzania – Mainland	2013	Africa 1	No	Lower-middle income
Uganda	2011	Africa 1	No	Lower-middle income
Not selected				
Guyana	*2002*	*Latin America and the Caribbean*	*No*	*Upper-middle income*
Kenya	*2005*	*Africa 1*	*No*	*Lower-middle income*
Tanzania – Zanzibar	*2013*	*Africa 1*	*No*	*Lower-middle income*
Zimbabwe	*2013*	*Africa 1*	*Yes*	*Lower-middle income*

Wave 1 sample-eligible partner countries, based on engagement with GPE 2025 Strategy and Operating Model and participation in other GPE evaluations

Country	Cohort	Cohort start date	Compact estimate date	Participation in other GPE evaluations	System Transformation Grant indicative allocation (US$ million)	Girls Education Accelerator (US$ million)	Multiplier allocation ceiling (US$ million)	System Capacity Grant ceiling (US$ million)
Cambodia	Cohort 2	01/10/2021	01/03/2023	Summative CLE 2019/2020	15.39		30.00	2.70
Congo, Democratic Republic of	Cohort 1 (pilot)	01/01/2021	15/09/2022	Prospective CLE; Formative COVID-19 Evaluation; Summative COVID-19 Evaluation	162.50	25.00	50.00	4.70
El Salvador	Cohort 1 (pilot)	01/01/2021	27/05/2022	Multiplier Evaluation	n/a	5.00	15.00	1.60
Nepal	Cohort 1 (pilot)	01/01/2021	25/07/2022	Prospective CLE	20.00		40.00	3.40
Sierra Leone	Cohort 2	01/10/2021	23/11/2022	Summative CLE 2018/2019	22.47		15.00	2.60
Tajikistan	Cohort 1 (pilot)	01/01/2021	04/05/2022	Summative CLE 2019/2020	10.00		15.00	2.60
Tanzania – Mainland*	Cohort 2	01/10/2021	05/07/2022		117.59		50.00	5.40
Uganda	Cohort 1 (pilot)	01/01/2021	15/09/2022	Summative CLE 2019/2020	162.50		50.00	3.80
Not selected								
Guyana	*Cohort 3*	*01/01/2022*	*14/11/2022*	*Summative CLE 2018/2019*	*5.00*		*5.00*	*1.50*
Kenya	*Cohort 1 (pilot)*	*01/01/2021*	*05/04/2022*	*Prospective CLE; Multiplier Evaluation*	*53.30*		*50.00*	*3.80*
*Tanzania – Zanzibar**	*Cohort 2*	*01/10/2021*	*10/10/2022*		*117.59*		*50.00*	*5.40*
Zimbabwe	*Cohort 2*	*01/10/2021*	*03/08/2022*	*Prospective CLE*	*24.83*	*12.42*	*30.00*	*3.30*

* Note: disaggregated figures for Tanzania – Mainland and Tanzania – Zanzibar were not available.

Wave 1 sample-eligible partner countries, based on selected characteristics of political structure and education system structure

Country	Federal vs unitary	Government effectiveness (2021)	Education spend as percentage of government budget	Percentage of all children in primary education	Percentage of all children in secondary education	Percentage of all children in early childhood education	Percentage of girls in primary education	Percentage of girls in secondary education
Cambodia	Unitary	−0.4226365	11%	92%	58%	70%	96%	63%
Congo, Democratic Republic of	Unitary	−1.7188781	20%					
El Salvador	Unitary	−0.3126774						
Nepal	Federal	−0.8715262	18%	120%	100%	89%	123%	101%
Sierra Leone	Unitary	−1.1145082	18%	87%	78%	42%	88%	78%
Tajikistan	Unitary	−0.5919222		95%	96%	13%	95%	94%
Tanzania – Mainland	Unitary	−0.6317798	22%	69%	33%	56%	72%	35%
Uganda	Unitary	−0.5702274	12%	53%	26%		54%	25%
Not selected								
Guyana	*Unitary*	*−0.2367316*	*18%*					
Kenya	*Unitary*	*−0.329994*	*24%*					
Tanzania – Zanzibar	*Unitary*	*−0.6317798*	*13%*	*69%*	*33%*	*56%*	*72%*	*35%*
Zimbabwe	*Unitary*	*−1.2429304*	*15%*	*90%*	*68%*	*57%*	*91%*	*67%*

Annex 4. Evaluation Report Outlines

Country-Level Case Study Report Outline (tentative)

Consistent structure of the case studies will be important to ensuring overall consistency between them, and efficiency when drawing on findings in a cross-country analysis for the synthesis report.

Title: GPE Country-Level and Thematic Evaluation – [Country name] Case Study – *descriptor*

1. Executive summary
2. Introduction
3. Country context
4. GPE in country
5. Theoretical framework, methods, evaluation questions and limitations
6. Country-level theory of change and transformative pathways (including assumptions)
7. Case study findings (subsections TBD based on each phase of the evaluation)

    7.1 Prioritized reforms and transformative potential

    7.2 GPE operating model and strategy and the identification and planning of the reform

    7.3 GPE operating model and strategy and the implementation of the reform

8. Conclusions

Annexes will include the evaluation matrix, data sources (documents reviewed, stakeholders interviewed etc.) and additional detail on subjects such as country context/political economy analysis, methodology, theory of change, social network analysis and others as appropriate.

Notes:

- The *descriptor* subtitle will be Baseline, Midline or Endline, as appropriate.
- The case study findings will be organized by Leading Evaluation Question. Those included above relate to Baseline evaluation questions.
- Recommendations will not be included for country-level case study reports. However, reports will include a set of strategic questions resulting from the case study, which may include gaps or views on further analysis required.

Synthesis Report Outline (tentative)

Title: GPE Country-Level and Thematic Evaluation – Phase XX Synthesis Report

1. Executive summary
2. Introduction and context

    2.1 Background

    2.2 Evaluation purpose and approach

    2.3 Methodology

    2.4 Structure of the report and color-coding scheme (*see notes*)

    2.5 GPE operational model and theory of change: progress and challenges

    2.6 Country contexts

    2.7 Limitations
3. Overarching observations and findings
4. Prioritized reforms and transformative potential
5. Baseline findings: GPE operating model and strategy and the identification and planning of the reform
6. GPE priority area: domestic financing
7. GPE priority area: gender
8. Conclusions
9. Recommendations

Annexes will include a list of any data sources accessed during preparation of the synthesis report (documents, additional interviews); and potentially additional supporting detail for the findings, in the interest of keeping the report itself concise.

Notes:

- The findings headings above reflect the Phase 1 synthesis report, which includes a section on the baseline findings from the first eight country-level case studies.

- The methodology will be the synthesis report methodology, not the generic methodology used for the country-level case studies.

- The annexes will also include a section summarizing how each of the remaining priority areas are featured in the sampled countries (such as relationship of priority area to chosen reform, status of priority area if featured in enabling factors and actions agreed to address gaps in enabling factors, and use of GPE grants/other mechanisms to address priority areas).

- A color-coding scheme (report section 2.4) will be used to report on the progress achieved in various evaluation questions and thematic areas. The scheme has been used in previous synthesis reports in GPE 2016–20. We suggest using it in order to simplify understanding of findings.

- The synthesis report findings will be organized by Leading Evaluation Question and will not necessarily address all the subquestions in equal amounts of detail. Findings against each will include a strength of evidence rating.

- The Phase 2 synthesis report will be in two halves: one synthesizing seven additional baseline country-level case studies (including analysis against gender equity and domestic financing themes) and also providing a comprehensive overview of the final baseline for all 15 partner countries; and one synthesizing the eight midline country-level case studies (focusing on the gender equity theme). A new section for the midline findings will be included, covering the GPE operating model and strategy and the implementation of the reform.

- The Phase 3 synthesis report will reflect the nature of the country-level case studies in Phase 3 which, as with Phase 2, will include two halves: one half examining seven midline country-level case studies (and the theme of gender equity) and also providing a comprehensive overview of the final midline for all 15 partner countries; and a summary of eight endline country-level case studies (examining both domestic financing and gender equity). A new section for the endline findings will be included, covering early signs of progress and results of the reform.

- The Phase 3 synthesis report will also include the following chapters, reflecting the summation of the evaluation:

  o Summary of key findings from Phases 1 and 2

  o Evaluation theory of change and underlying set of assumptions – evidence of its viability, strengths and weaknesses and explanatory factors for this.

Annex 5. Communications Protocols

Body	Agreed processes	Mode of communication	Frequency
1. Internal – core management team	Regularly scheduled meetings with standing agenda to discuss evaluation progress (through review of Microsoft Planner), review of issues raised (review of issues log) and making key decisions on evaluation design, methodology and personnel	Online meetings scheduled, chaired by the Evaluation Manager	Bi-weekly
		Regular contact via email and dedicated Microsoft Teams channel	Ad-hoc
2. Internal - core evaluation team	Regularly scheduled meetings with standing agenda to discuss evaluation progress (including country-level case studies through review of Microsoft Planner, and on thematic analysis, synthesis and learning activities), review of risks (risk matrix) and review of issues (review of, add to issues log) They will identify issues that need to be escalated to the core management team (including changes in risks, newly identified issues, changes to workplan or timelines and changes to personnel) They will also discuss the operationalization of key decisions made by core management team and learnings from across studies (e.g., cascade to country-level case study teams)	Online meetings scheduled, chaired by the Evaluation Manager	Weekly
		Regular contact via email and dedicated Microsoft Teams channel	Ad-hoc
3. Internal – country-level case study teams	Regularly scheduled meetings with standing agenda to discuss ongoing technical activities, country-level case study progress (reporting to Microsoft Planner) and identification of issues (add to issues log). These meetings will be held for the duration of when country-level case studies are in operation	Online meetings scheduled. Chaired by Case Study Lead (or Qualitative Evaluator – to be determined for each country-level case study)	Weekly
		Regular contact via email and dedicated Microsoft Teams channel	Ad-hoc
4. R&P team	Formal meetings with a pre-agreed agenda to discuss operational and technical issues and decision-making points in structured meetings with the R&P team	Online meetings scheduled. Chaired by the Evaluation Manager or another member of the core management team. Notes taken by the Evaluation Manager	Monthly

Body	Agreed processes	Mode of communication	Frequency
	Regularly scheduled quick catch-ups to share latest progress updates	Online meetings with the Evaluation Manager and Anne Guison-Dowdy (and other select members of the Evaluation or R&P team as needed)	Weekly or as needed
	Any other informal update	Email from Evaluation Manager or core management team	Ad-hoc
	To raise issues that have implications for the budget or compliance	Raised directly by Project Director or Evaluation Manager with the R&P team via email	Ad-hoc
5. Other GPE Secretariat teams/staff	Engagement with Country Team Leads (CTLs) for country-level case studies for initial consultation, to conduct stakeholder mapping and identify any data gaps as part of the country case study work	Where initial communication with CTL is required, any introductions should be first facilitated by R&P Once contact has been established, on new issues, the Evaluation Manager and Anne Guison-Dowdy should be included in email copy in the first instance	At start of country-level case study
	Informal meetings with the Learning Leadership Team or Knowledge and Innovation Exchange Team to identify potential synergy or opportunities in order to curate knowledge exchange, aimed at ensuring findings are not shared at the end of the process but as they emerge, feeding back into program design in real time	Online meetings scheduled by the Evaluation Learning Team, with Anne Guison-Dowdy and the Evaluation Manager included in the email	Ad-hoc
	Contact with further GPE Secretariat Staff (such as CEP/ Education Policy and Learning/Finance and Grant Operations/Partnership team members) either for consultation (such as on thematic areas, grants and civil society engagement) or for global interviews	Facilitated by R&P only or, where direct contact with the evaluation team is made, R&P will be made aware of the communications and included in the discussion where relevant	Ad-hoc and as part of global-level data collection
6. In-country stakeholders	For initial country-level stakeholder engagement, CTLs (see above) will put Case Study Teams in touch with key in-country contacts (including ministries of education focal points, Coordinating Agencies and Grant Agent Representatives) for country-level interviews, country-level stakeholder surveys,	Initial email introduction to be provided by CTL (via R&P)	For country-level case study planning as needed

Body	Agreed processes	Mode of communication	Frequency
	country-level debrief workshops, follow-up clarifications via email and learning events	Remote initial consultations with selected stakeholders (if available), including on theory of change In-person or remote country-level interviews In-person or online survey In-person or remote country-level debrief workshop	As part of country-level data collection
	Grant Agents and/or Coordinating Agencies are likely conduits to further in-country stakeholders, which may include representatives and policy makers from ministries of education and other line ministries, other local education group members including civil society groups, teacher organizations, donors and technical partners, as well as – if deemed appropriate – other stakeholders that are not part of the local education group for country-level interviews, country-level stakeholder surveys, country-level debrief workshops, follow-up clarifications via email and learning events	Initial email introduction to be provided by Grant Agent, Coordinating Agencies etc. In-person or remote theory of change workshop (for select, relevant stakeholders as identified in the stakeholder mapping process) In-person or remote country-level interviews In-person or online survey In-person or remote country-level debrief workshop	For country-level case study planning as needed As part of country-level data collection

Annex 6. Risk Management Approach

The implementation of this evaluation will require robust risk monitoring and management. We have put in place a risk management approach that guides how the evaluation team will prevent issues from occurring, minimize the level risk and avoid transferring risk to study participants or to GPE.

Our risk management approach comprises the following three steps:

1. **Risk identification and analysis.** All members of the evaluation team have contributed to risk identification and analysis, applying their different areas of expertise and experience, contextual knowledge as well as intelligence from the client to ensure that all types and aspects of risk are considered. The risks we initially identify relate to the external context, operations, methodology, delivery, safeguarding, data management, sexual exploitation, abuse or harassment (SEAH) or conflict of interest. For each risk analyzed, we assess the likelihood of each identified risk occurring and its impact on the work if it did occur. Contextual risk will vary with geography, and so will be assessed for each location where evaluation activities are planned as part of our pre-departure risk assessment. Mitigation actions are clearly defined and assigned to an appropriate team to ensure ownership, with support and oversight from the Evaluation Manager. The risk analysis and mitigation actions are documented in a risk matrix.

2. **Management and mitigation.** We will plan and undertake mitigation actions for each identified risk. Some risks will be tolerated; others will be treated (their likelihood or impact reduced) or terminated (avoided through a change in approach, e.g., related to geography).

3. **Monitoring and adjustment.** Each aspect of the risk matrix (definition, analysis and mitigation plan) will be formally reviewed by the Evaluation Manager regularly. In core evaluation team meetings, we will undertake active bi-weekly monitoring of risks and all team members will raise concerns about risks which are escalating (increasing in likelihood and impact) as soon as they arise, so they can be acted upon without delay. Where new or escalating risks are identified (e.g., risks that are increasing in likelihood and/or impact), this will be brought to the attention of the core management team either through email or in regular meetings and communicated to country-level case study teams as appropriate. Where the core management team deems relevant, escalating risks and suggested responses will be raised with R&P.

In addition, Triple Line has policies in place which apply to all contracts we implement and to all partners and staff we work with. The table below lists our relevant policies that cover different types of risk.

Risk types	Triple Line policy
External context	Duty of Care policy Triple Line Duty of Care Policy.pdf
Conflict of interest; Delivery; Safeguards	Code of Conduct Triple Line Code of Conduct.pdf
SEAH; Delivery; Safeguards	Code of Conduct Child Protection Policy Safeguarding Policy Reporting Concerns (whistleblowing)

	Triple Line Safeguarding Policy.p / Triple Line Child Protection Policy 2019 / Triple Line Whistleblowing & Rep
Data management; Methodology; Safeguards	Data Protection & Cyber Security Policy Triple Line_Data Protection and Securi

Annex 7. Risk Matrix

Key: L = Low, M = Medium, H = High

Type	Risk	Likelihood	Impact	Mitigation	Who owns it?
External Context	1. Political (or other forms of) insecurity (e.g., acts of terrorism, armed conflict, crime or civil unrest) or natural disaster/manmade hazards reduces the availability of key informants to participate in case studies	L	M-H	We considered this as part of our sampling criteria, excluding potential cases where respondents may not be available for remote interview due to political or contextual factors. We will flag to the R&P team at the earliest convenience when problems emerge (e.g., rapid changes to the context of any of the selected case study countries) in order to allow remedial steps (e.g., reduction in scope where possible).	Core management team (country-level sampling and methodological adaptations)
External Context	2. Changes in political/security situation, conflict, natural disaster, weather or COVID-19 puts team members and/or stakeholders at risk; inhibits travel and reduces accessibility	L	M-H	We will closely coordinate with In-Country Researchers prior to fieldwork in order to monitor any changing contexts. Recognizing the particular risks that can arise in lower-income countries in which GPE operates, duty of care policies are in place and a specialist contractor (International SOS) is employed by Triple Line (through its parent company, IPE Global) to assess and manage security risks for personnel in specific locations and to mobilize/support intervention on the ground should risk materialize. As necessary, we will adjust the evaluation approach to minimize these risks, e.g., by rescheduling data collection, considering alternative locations and employing remote/virtual means of data collection. Evaluation teams will also be sensitive to additional constraints on stakeholders' time and work flexibly to ensure that interviews are conducted succinctly and when convenient.	Core evaluation team (country-level case study roll-out) Evaluation Manager (enforcement of policies)
Operational	3. International and national team members cannot obtain government authorization to implement evaluation activities	L	M	We will work closely both with our In-Country Researchers and with key evaluation stakeholders in-country (such as ministries of education representatives, Grant Agents and Coordinating Agencies) to understand and anticipate any requirements for authorization for evaluation activities. Research activities can also be shifted to our in-country researcher who will have a strong presence in the relevant country and be known and respected by local stakeholders including government. Where possible, we will identify Case Study Leads who are based in the country or region.	Case study teams

Methodology	4.	Low levels of stakeholder engagement or participation at the country level	M	M-H	We will engage with key GPE Secretariat team members (such as CTLs) to develop a stakeholder consultation strategy at the start of each country-level case study, in order to identify the most relevant respondents and to understand any restrictions to access or dynamics that might create reluctance to engage with the evaluation. The strategy will also identify the most effective route for engagement, such as building on existing mechanisms or relationships, to determine how best to approach respondents to participate in the case study. This may involve identifying the most appropriate route for introductions or method for interview (remote, in person, etc.). At the start of any respondent engagement, we will ensure that the purpose and questions of the interview are articulated proactively. We will also streamline communications and minimize touch points in order to prevent research fatigue.	Case study team
Methodology	5.	Slower than anticipated progress in the development of partnership compact	M	M-H	Although sampling will consider the date for expected completion of country partnership compacts and exclude any beyond a cut-off point, in order to ensure that this process is completed prior to data collection, there could be subsequent slippage in expected completion dates. We will keep in close contact with the R&P team to monitor any changes to the expected completion dates, in order to plan mitigation. This may include slightly delaying data collection (e.g., staggering the roll-out of country studies during a phase). Where a country is close to compact completion, extra data collection (e.g., additional interviews) may be used prior to the start of in-country fieldwork to ensure that the country-level case study team can gather key information normally included in the partnership compact.	Core management team (country-level sampling) Core evaluation team (country-level case study roll-out)
Methodology	6.	Low response rate from the survey questionnaire	H	M-H	We propose to administer the survey questionnaire "live" during data collection visits to ensure input from key stakeholders. We will avoid bias by ensuring respondents complete it in private after having explained the process to them. We will keep it short and simple to maximize the response rate.	Core evaluation team (survey design) Case study team (survey administration)
Methodology	7.	Lack of consistency or coherence between country-level case studies	L	M-H	We devised a QA framework to ensure consistency and coherence between country-level case studies. The QA framework includes provisions to support both methodological consistency (including fidelity and appropriate adaptation from the core evaluation design) as well as stylistic and presentational consistency.	Core evaluation team (with support from core management team to undertake

						some QA activities)
Methodology	8.	Missing data means that the evidence base may be insufficient to properly address some evaluation questions	L	L	We will ensure that there is strong engagement with the R&P team from the start, setting the tone for close consultation and fit between evaluation design and evaluation needs. We will collaborate with the GPE Secretariat (including R&P and CTLs) to help ensure that the evaluation team engages with the right people (audience) at the right time. We will also provide regular updates and feedback to R&P to discuss data collection progress and develop mitigation plans when progress is off track. Using either the staggered approach to data collection (within a phase) or our phased approach (between phases), we will adapt the methodology as required to address critical gaps. All country-level case study reports will include a strength of evidence assessment to transparently report on the validity of evaluation findings with regards to gaps in the evidence base. We note that availability of data to support reform enactment – some of which might also be needed for the case study – is an assumption (no. 10) which will be tested through the case studies.	Core management team (evaluation design) Core evaluation team/Case study teams (on reporting)
Delivery	9.	Genuine learning (generating and reflecting on evidence and applying lessons) is reduced due to time pressures or mismatched timing between decision-making points and delivery of findings	L	H	We will work closely with R&P to understand key decision-making points (and timing) and the needs of evaluation stakeholders in the development of the workplan for each phase. We will ensure there are check-in points throughout the evaluation to revisit and realign on the learning strategy as needed, based on any changes to timeline or contexts for learning opportunities, including with the GPE Secretariat's newly formed 2LT.	Evaluation Learning Lead and Learning Manager
Delivery	10.	Delays to delivery of complex workplan required for multiple case studies per phase (with dependencies between case studies and synthesis report) and QA process at all stages; budget over-runs for the evaluation team	M-H	M	In Phase 1, we will stagger the delivery of the country-level case studies, carefully plan and closely monitor core and country-level case study teams inputs, including providing sufficient time for QA, on an ongoing basis. This will allow us to monitor the evaluation's initial roll-out as a whole with regard to resource use and workplan timing. We will feed learning about evaluation roll-out from each phase into planning for the next. We will flag to the R&P team at the earliest convenience to discuss if problems emerge in order to allow remedial steps (e.g., reduction in scope where possible).	Core evaluation team (in conjunction with case study teams)

Delivery	11. Evaluation team members are not available when needed or due to illness, accident or bereavement or leave the team	L	M	We will ensure that all team members recruited to the core or country-level case study teams are available for and committed to the duration of the evaluation component they are engaged on. Core evaluation team members will also work across different areas (including across country-level case studies) to ensure that study knowledge is spread throughout teams (and not individuals). However, personal situations may change for team members. Where required, any changes in key team members will be discussed with GPE prior to replacement. To replace team members, we will first seek replacement inputs from within the evaluation team to support a smooth transition and knowledge retention and to enable an immediate response. For longer term arrangements, or where back-filling within the evaluation team is not possible or sustainable, we will look within our consortium members for similarly qualified or experienced staff or, where not available, to trusted associates. All new team members will undergo standardized evaluation on-boarding and appropriate checks for due diligence and conflict of interest.	Core management team (personnel decisions)
Safeguarding	12. Evaluation activity causes unintentional harm or risk to participants	L	H	Our stakeholder consultation strategy will identify any risks for evaluation participants that might occur as a result of participation in our evaluation. Where participation offers risk, we seek to either mitigate this risk (e.g., if physical safety is a risk, we may opt for remote methods for data collection); where the risk cannot be mitigated, data will not be collected. Prior to participation in our evaluation, we will obtain informed consent to ensure respondents understand the purpose of the evaluation and that they are entitled to withdraw at any point. All evaluation team members will undergo safeguarding and ethical research training as part of on-boarding and all evaluation team members are subject to Triple Line's safeguarding and ethical research policies.	Case study teams (consent processes) Core evaluation team (providing training) Evaluation Manager (enforcement of policies)
Safeguarding	13. An evaluation team member's personal safety is compromised during fieldwork or travel	L	H	We will ensure that a country-level risk assessment is conducted prior to any travel and field data collection (to support both traveling team members and researchers based in-country). Triple Line has robust duty of care policies and procedures, including mitigation, escalation and emergency procedures.	Evaluation Manager (enforcement of policies)

Safeguarding	14.	Mistrust on the part of target interviewees who are reluctant and fearful of engaging or put at risk through the evaluation	L	M-H	We have an ethical research policy (on which all evaluation team members will be trained) to guide our research conduct to ensure that no participant feels under pressure to be interviewed. We will not pressure individuals or groups to engage and will recognize that reluctance to speak is a finding in and of itself. Where relevant, we will include provisions to ensure that data collection is conducted in a gender-sensitive, inclusive and culturally sensitive manner (such as with female researchers and using local languages).	Case study teams Core evaluation team (managing case study team resources)
SEAH	15.	Participants in evaluation activities, including the evaluation team members and particularly vulnerable groups, are subject to potential harm related to SEAH This could be due to lack of awareness or permissive cultural norms, context-related factors which indicate heightened risk of SEAH or the exploitation of differential power relations	L	H	The consortium has policies and procedures in place to safeguard against SEAH internally and externally. These include the following: • Due diligence will be carried out in the recruitment of all team members. Enhanced due diligence, e.g., Disclosure and Barring Service (DBS) check, carried out where engagement with vulnerable individuals is expected. Evaluations will not bring third parties into contact with vulnerable individuals (e.g., in FGDs) where there are power imbalances that could be exploited. • All evaluation team members will be provided with training on Triple Line's safeguarding and SEAH policies as well as all safeguarding reporting procedures. • We have a safeguarding reporting process. Triple Line has appointed two point-persons to whom complaints can be lodged and who are trained to provide victim-centered support. We have a register for complaints and a mechanism to track complaints until they are resolved. Point-persons report to the Triple Line Board and one has a position on the board.	Evaluation Manager (enforcement of policies) Core evaluation team (providing training) Triple Line Safeguarding Leads
Data management	16.	Data stored or collected for the evaluation are inappropriately accessed	L-M	M	All data for the evaluation, including primary data collected and GPE data and documents used as part of the evaluation, will remain securely stored on privately accessed SharePoint sites of GPE or Triple Line. All access to SharePoint sites is granted only through company-registered Microsoft 365 logins which use two-factor authentication. All evaluation team members are required to abide by Triple Line's Data Protection and Cyber Security Policy and will not download the data. Access to data is restricted to a need-to-know basis for team members.	Evaluation Manager (enforcement of policies) Core evaluation team (providing training)

Data management	17. Access to sensitive data puts individuals at risk	L	M-H	We will put in place data protection and minimization processes where people's security would be in danger if sensitive data are accessed.	Evaluation Manager (enforcement of policies)
				A knowledge management system will be established to log and register data systematically. Our SharePoint site contains specific permission settings to ensure restricted access to sensitive data.	
Conflict of interest	18. Evaluation judgment is biased or compromised due to individual conflict of interest of team member(s)	L-M	M-H	We have strict conflict of interest processes in place for all members of the team and consortium partners. We will conduct conflict of interest checks with any personnel before recruitment to the evaluation team.	Core management team (personnel decisions)
				Any evaluation judgments made will be checked for quality and bias by the QA team, which includes team members from different members of the consortium.	Core management team (QA function)
				For each evaluation, any concerns over conflict of interest will be logged in our issues register and regularly examined.	

Annex 8. Research Guidelines

Introduction

This document sets out the main standards and processes for developing and applying research guidelines that will be used to support delivery of the Country-Level and Thematic Evaluation, implemented by Triple Line, Technopolis and Learn More.

We first set out the key principles that guide this evaluation's conduct, which are based on GPE's Evaluation Policy. We then set out the further processes that will be utilized throughout the evaluation, with a focus on the whole chain of research activities (e.g., fieldwork planning, data collection and data storage).

We also include as part of this annex copies of the relevant Triple Line compliance documents that apply to all individuals working on the evaluation team more broadly (including the core management team, core evaluation team and country-level case study teams) and a copy of the data collection consent form which will be shared with all research participants as part of this evaluation.

Key Principles

The basis for the key principles that will guide this evaluation is the principles set out by the GPE Evaluation Policy. Below, we set out the ways in which we have considered GPE's evaluation principles and how we will respond to them throughout the evaluation.

Principle	Our conduct
Independence and impartiality	Our evaluation team has full independence from GPE. We have no vested interest in any particular area of the evaluation or in the findings. Each of our team members has been vetted for any conflict of interest. While the ultimate accountability of the evaluation falls with our evaluation Project Director, the evaluation is led by a core team comprising members from all three consortium members and one member external to all three organizations, in order to mitigate against any bias. We work closely with GPE's R&P team, who facilitate full access to GPE reporting information and allow for full autonomy in carrying out data collection, analysis and reporting.
Credibility	To ensure the credibility of our evaluation, we have assembled a team that brings together expertise in a number of critical aspects for the evaluation, including on complex evaluation methodologies and education system transformation. Our inception phase has included extensive consultations across GPE to ensure that we have fully understood the GPE 2025 operating model and that we have selected the most appropriate evaluation methodologies. The credibility of the evaluation will be further tested through a review of the evaluation's design by key evaluation stakeholders such as GPE's R&P team and the independent technical advisors assigned to this evaluation, and subsequent revisions and updates by the evaluation team. Furthermore, there will be ongoing engagement with R&P throughout the evaluation as well as reviews built into the evaluation plan to allow us to reflect on the evaluation's progress and results and to revise the evaluation approach and methodologies in order to support its continued relevance.
Transparency	We uphold the principle of transparency by working with R&P to ensure that aspects of our evaluation design and process are publicly available, including this inception report and detailed description of the evaluation process once completed, as part of the evaluation reports.

Utilization for learning	Our evaluation is participatory and utilization focused. We will work closely with the GPE Secretariat to co-design each step of this evaluation to ensure that the evaluation is timed to support key decision-making moments both within the GPE Secretariat and for the GPE Board. We will also ensure that the evaluation activities are socialized at the country-level in the countries in which evaluation activities will take place in order to support country-level stakeholders, such as members of the local education group and notably the ministries of education, to engage with our evaluation findings to support their needs.
Ethical Principles	Our evaluation adheres to the highest ethical principles for evaluation. The consortium partners are value-driven organizations committed to high ethical standards for their work.

• **Trust and respect** will be built through engagement with key evaluation stakeholders and participants along the evaluation process. By working with and through the GPE Secretariat team, we aim to ensure that principles of respect and recognition are followed and trust is built. We will always maintain respect for distinct perspectives and the cultural and contextual circumstance.

• **Principles of human rights** will be applied during data collection to ensure no harm is done to any participant engaged through the evaluation. We will also aim to ensure that all aspects of our methodology (e.g., consultation with end users during case studies) consider inclusivity across a range of dimensions, including social, economic and power.

• **Voluntary participation and informed consent** are ensured for all interviewees and focus group participants who take part in the evaluation. Participation in the evaluation is voluntary and participants will be informed that they are free to leave the process at any time. All evaluation participants will be required to give their consent (verbal or written) to participate.

• **Confidentiality and privacy** will be ensured and respected with all primary data collection. When findings are presented, all evaluation participants will be anonymized, so that as much as possible, specific findings and comments cannot be traced back to individuals. Data will be stored systematically and securely so that they are available and clearly accessible to the evaluation team and, with appropriate anonymization, to GPE as required. Arrangements will be made to store or destroy data as appropriate after conclusion of the contract.

• **High standards for safeguarding** will be applied throughout the evaluation. All evaluation team members will be required to adhere to Triple Line's Code of Conduct and Safeguarding Policies. Risks on key safeguarding concerns, such as regarding SEAH, have been identified for the evaluation and our risk management policies and risk matrix outline our mitigation strategies. |
| Participation of end users | We will ensure that diverse and distinct views are included as part of our evaluation. The use of country-level stakeholder mapping, a political economy approach and in-person fieldwork ensure that our evaluation considers the power dynamics embedded in education systems and offers both detailed thinking and flexibility to ensure that diverse views are included in the evaluation. For country-level studies, we will work with and through country-level partnership structures, such as local education groups, for consultation in the design of country-level case studies as well as to disseminate findings. |
| Gender equality and inclusion | All evaluation questions are informed by a gender and social inclusion lens and disaggregated data are sought and collected as far as is possible. The perspective of different stakeholders, especially in accessing education will be actively |

> measured by a carefully balanced team of women and men of different cultures and ethnicities.

Research Processes

Triple Line recognizes the importance of robust policies, processes and procedures to ensure the smooth running of in-country research.

At the start of each country-level case study, we will develop an "on-boarding" pack with relevant documentation (such as related to due diligence and conflict of interest, essential policies and duty of care – see Annex 6) to ensure all country-level case study team members who will be engaged in the delivery of country-level case studies have all the necessary knowledge and capabilities required to conduct safe, inclusive and equitable research in the specific country to which they are assigned. The core evaluation team will on-board country-level case study team members on the relevant materials in a virtual workshop, in order to ensure their commitment to follow Triple Line's policies, processes and procedures (included below). They will be required to provide a written signature.

Safeguarding Policy	Triple Line Safeguarding Policy.p	Data Protection and Cyber Security Policy	Triple Line_Data Protection and Securit
Duty of Care	Triple Line Duty of Care Policy.pdf	Code of Conduct	Triple Line Code of Conduct.pdf
Child Protection Policy	Triple Line Child Protection Policy 2019	Whistleblowing	Triple Line Whistleblowing & Rep
Modern Day Slavery	Triple Line Modern Slavery Statement 202	Media Policy	Triple Line Media Policy.pdf
Research Ethics	Triple Line Research Ethics Policy 2023.pc	These documents are available on request.	

Due Diligence and Ethical Conduct

For all our subcontractors, we will ensure that there are due diligence and conflict of interest processes in place to guarantee that they will demonstrate ethical conduct and be able to conduct in-country research effectively as part of this assignment. All our evaluation activities will adhere to the ethical provisions in **Triple Line's Research Ethics Policy**.

Our core principles for ethical research follow:

- Research should aim to maximize the benefit for individuals and society and minimize risk and harm
- Rights and dignity of individuals and groups should be respected
- Participation must be voluntary and appropriately informed
- Research should be conducted with integrity and transparency
- Lines of responsibility and accountability should be clearly defined
- Independence of research should be maintained and where conflicts of interest cannot be avoided, they should be made explicit.

We will ensure that ethical issues are considered across the lifecycle of the assignment, including processes for reporting ethical concerns, and adjusting research to minimize risks to participants.

Protection From Harm

We will ensure that data collection does not put research participants or country research teams at risk, for instance by consulting with the R&P team and other GPE stakeholders as appropriate to understand any harm which could arise in specific contexts and planning to avoid this. Special measures to ensure the safety of participants might include providing alternative ways to participate, arranging for follow-up interviews in a different location or time, or offering accessible versions of background project information. Refer to our risk matrix for more information.

Overall, our **Code of Conduct** includes provision to **adopt a rigorous stance on all human rights abuse, including sexual exploitation and abuse, modern slavery and human trafficking.**

This includes the commitment to the following:

- Promote a "zero tolerance" approach to all forms of human rights infringements including sexual exploitation and abuse, sexual harassment, modern slavery and human trafficking

- Immediately report to the relevant authority any known act of human rights abuse, including sexual exploitation and abuse, modern slavery or human trafficking committed by its employees, contractors, partnering firms or any third-party agent in the course of their commission on any lead, bid or project

- Suspend any employee or supplier suspected of partaking/aiding/abetting/failing to recognize or notify Triple Line of any forms of human rights abuse, including sexual exploitation and abuse, modern slavery and/or human trafficking, and subsequently terminating their contract should any later investigation find they have acted in direct contravention of our "Code of Conduct."

Our guidance on **Reporting Concerns (Whistleblowing)** sets out Triple Line's procedure for dealing with all reports of behavior that we may believe to be in breach of Triple Line's policies, including safeguarding against sexual abuse and exploitation, bullying and harassment, including sexual harassment, Code of Conduct, child protection and modern slavery.

Safeguarding

Furthermore, Triple Line has stringent safeguarding and child protection policies in place and the Project Director will ensure compliance. We will ensure that country-level case study teams, including In-Country Researchers, are aware of their safeguarding responsibilities.

Our **Safeguarding Policy** sets out measures for ensuring everyone, including children and at-risk adults, are protected from harm that arises from interacting with our staff, team members or research activities. This includes responsibilities concerning child safeguarding, adult safeguarding and protection from sexual exploitation and abuse.

Child Protection

In particular, our **Child Protection Policy** sets out our professional duty to ensure that the children we work with are safe from harm and that any concerns that emerge through our work are addressed. This includes ensuring the following:

- In addition to the security checks that all Triple Line employees and subcontractors are subject to upon recruitment (which meet His Majesty's Government Baseline Personnel Security Standard), all employees who work with children have current enhanced DBS checks.

- All research participants are informed about who to contact if they have any concerns about Triple Line employees or subcontractors.

It also sets out procedures for responding to and reporting a child protection issue, should one be encountered.

Prevention of Sexual Exploitation, Abuse and Harassment

Finally, we note that GPE has a zero-tolerance policy against all forms of SEAH in line with its SEAH policy. As described above, Triple Line has the following policies in place to manage SEAH risks, including risks associated with direct contact with children and other stakeholders at the community level: (a) Code of Conduct; (b) Safeguarding Policy; (c) Child Protection Policy; and (d) Guidance on Reporting Concerns (Whistleblowing).

In our risk matrix, we have identified potential SEAH risks and mitigation measures.

Duty of Care

We also have robust standards in place for our duty of care commitments. The work and travel of Triple Line staff and subcontractors will be governed by Triple Line's **Duty of Care Policy**, in addition to specific provisions tailored to the particular requirements of the evaluation and countries to be visited. The key components of our duty of care approach follow:

- **Risk assessment.** Prior to any staff deployment, a more detailed risk assessment will be conducted by our core evaluation team. Risk assessments will be based on the latest information available from local sources, including from Foreign, Commonwealth & Development Office (FCDO) HQ and FCDO country offices, and country-level partners in consultation with International SOS; and will inform risk management plans tailored to the risk environment in each country to be visited. This will also include the following:

  - A security risk assessment outlining the threats and potential risks that may affect the country-level case study team members and assets while performing under the contract

  - A brief statement of anticipated contingency plans and mitigation measures (areas for consideration include updates to the security risk assessment, personnel security briefings, physical security measures, medical care arrangements and evacuation plans).

- **Pre-deployment preparation.** Prior to deployment, we will ensure that all traveling evaluation team members are fully briefed and aware of the risk context, and that they have the required training appropriate to the location, its security context and in-country locations. We monitor whereabouts and wellbeing over the course of the mission.

- **Live/ongoing risk monitoring.** Through International SOS, we have access to reliable mechanisms to monitor risk on a live/ongoing basis. Key security actions required will be documented, and all country-level case study team members briefed on these. The risk matrix will be updated on a regular basis (as appropriate to the location) with activity-specific risk assessments conducted at the time of field activities, as specified above.

- **Personal equipment.** All staff are provided with suitable equipment according to their role and tasks, and the geopolitical context to which they are being mobilized. In particular, we ensure that staff are equipped with appropriate communications equipment (including radio and/or satellite phone where relevant), as well as protective personal equipment based on travel advice and our own risk assessment. We verify that all equipment is fully operational prior to deployment, and through regular routine checks for longer deployments.

- **Emergency response systems**. The risk analysis will inform tailored emergency response plans, prepared in consultation with in-country experts and advisors as well as our retained security provider. These will cover a range of potential scenarios (including deteriorating security, terrorism, kidnapping and medical emergency) and will include evacuation plans. The plans will identify key actors and responsibilities, establish emergency communications plans and define processes for ensuring plans remain actionable and equipment and technology are functional.

For our subcontractors, we will also require sensitive information such as IDs/passports and any information on medical conditions as part of our due diligence process. This information is required for awareness, in case assistance is required in the case of emergency, as well as for insurance purposes. We will follow the European Union's General Data Protection Regulation (GDPR) when storing private and confidential information on SharePoint.

The core evaluation team will use a checklist spreadsheet tool to monitor tasks completed pre-departure across the consortium.

Principles for Data Collection Processes

Ethical data management is a fundamental duty in monitoring and evaluating interventions in the humanitarian setting, especially when it involves children. Our approach is to be "open as possible, closed as necessary" and GDPR is integrated into all use of data. All efforts will be made to ensure any primary data collected are disaggregated by gender and other intersecting characteristics, as relevant. Key principles of our data collection processes follow:

- **Data minimization:** we will gather additional data only where these demonstrably contribute to the overall assignment and where it can be ensured the process does not put people at risk.

- **Privacy and confidentiality:** informants will receive a clear commitment to confidentiality and explanation that all data will be anonymized, including the processes we will undertake to do so. All data will be stored and de-identified (e.g., removed of any identifying features and identifying features will be secured using a password-protected key) and stored within Triple Line's secured SharePoint site. Further data security precautions, such as full anonymization, will be undertaken to ensure the privacy and wellbeing of particularly vulnerable groups.

- **Informed and voluntary consent:** we will ensure that all participants in our evaluation understand the purpose of any research and how data will be used. We will emphasize that participation is voluntary, can be withdrawn at any point and ensure that consent is freely given before starting any interview. Processes for this will be included as part of our data collection tools.

- Although unlikely to be required for this evaluation, all data collection involving children will involve both **consent (from parents or guardians) and individual assent** and be conducted in the presence of a third party or guardian and by a researcher who has undergone the relevant security checks. More information on this is set out in Triple Line's **Child Protection Policy**.

- **Respect for cultural sensitivities and human rights:** we will ensure that data are collected with respect for cultural sensitivities and human rights, which includes ensuring that data are collected in an appropriate language. We will also ensure that data collection does not put research participants or researchers at risk. We will consult with relevant country-level stakeholders (including CTLs and Grant Agents) and our In-Country Researchers to understand any harm which could arise in specific data collection contexts and create plans to avoid this. Special measures to ensure the safety of participants might include providing alternative ways to participate, arranging for follow-up interviews in a different location or time, or offering accessible versions of background project information.

Data Security

All members of the evaluation team will abide by Triple Line's overarching **Data Protection & Cyber Security Policy**. This policy protects all data shared by the client and collected in the course of the assignment, and was developed in accordance with the UK Data Protection Act (2018) which is compliant with GDPR. Key provisions in our policy relevant to these evaluations follow:

- **Section 10 – Handling of Project Data.** All parties with access to project data are responsible for ensuring the implementation of systems to deter, detect and resist intrusion and unauthorized access to data. This includes encryption control for data that are stored as well as data that are shared or transmitted.

- **Section 11 – Handling Sensitive and Confidential Client Data.** For data identified as confidential, higher levels of security may be implemented, such as encrypting documents, using watermarks, password-protecting or restrictions around emailing, printing and using cloud-based services.

Our policy and website also detail instructions for relevant parties to exercise their right to access personal data records. If any such request is received in relation to data collected or held for the undertaking of this assignment, we will also inform the client of the data request and collaborate, if necessary, in preparing the appropriate disclosure.

Data Management

Program documentation shared by GPE will be retained in the same structure, to retain context. Our approach to data management will be to index this corpus, assigning unique codes to each document for internal usage withing the consortium – including document review, coding software for qualitative analysis and management of processes.

Data Storage and Sharing

Under our data protection policy, project data will not be stored, copied or shared by any team member other than as far as is necessary for providing the services and obligations under the agreed service to GPE. The sharing of information with third parties and experts engaged from outside the consortium will follow a need-to-know principle and depend on the sensitivity of information. All parties with access to project data are responsible for ensuring the implementation of systems to deter, detect and resist intrusion and unauthorized access to data. A number of mechanisms to do this can be found within our **Data Protection and Security Policy**.

At the close of the evaluation, any information that will not be made freely available in the public domain will be destroyed in a way that prevents reconstitution.

If survey data are shared beyond the immediate team conducting the analysis, survey responses will be anonymized through the removal of data fields that can be used to identify individuals. Such identifying data will also only be collected for the purpose of following up with respondents in the case of unclear responses (the surveys are likely to be purposive samples of expert individuals, rather than a randomized sample).

Access to Data

As these evaluations will involve the mobilization of country-specific evaluation teams and external experts outside the salaried staff of consortium members, we will also manage internal access to project data (including documents shared by the client) according to necessity and relevance. External country evaluation team members will have limited access to our document repository, with permission to access only country-specific documents and a curated selection of program-wide documentation.

Interview Consent Form

What is the purpose of the interview?

You have been invited to take part in an interview as part of the 2022–26 *Global Partnership for Education* (GPE) Country-Level and Thematic Evaluation, conducted by *Triple Line Consulting*, *Technopolis Group* and *Learn More* (who form the evaluation team).

The purpose of the Country-Level and Thematic Evaluation is to evaluate and generate continuous learning on GPE's global and country-level support as part of GPE's 2025 strategy and operating model. The objectives of the evaluation are to understand the relevance, coherence, efficiency and effectiveness of GPE's global and country-level support for countries' system transformation efforts and with respect to GPE's key priority areas. We will also seek to identify promising practices and formulate recommendations to support GPE to strengthen its support to partner countries.

As part of this evaluation, we are inviting key stakeholders to take part in interviews who have been involved in or have knowledge of GPE's operating model and strategy.

What will happen in the interview?

Our evaluation conforms to *GPE's Evaluation Policy*, **which requires that GPE-related evaluations be conducted with adherence to the highest ethical principles.**

Prior to the interview, your interviewer will confirm the time (and location or link to call) as well as let you know how long the interview will take.

Ahead of the interview, the interviewer will introduce the evaluation team taking part in the call, which may include translators where necessary. The interviewer may also ask permission to record the interview with an audio-recorder and to take notes. You may also request that the interviewer (a) does not record your name or (b) records your name but does not include it in any report (e.g., in a list of interviewees).

What will we do with the data from the interview?

The data from this interview will only be used for this evaluation. If you permit, we will disclose your participation in our evaluation in the report, but any answers you provide will be treated confidentially and any information (such as quotations) we present in our evaluation reports will be anonymized.

How are the data stored?

We keep any data collected for our evaluation in a secured corporate system. All data stored for the evaluation will be de-identified, meaning it will not be possible to identify you from your answers. Data from the evaluation will be securely stored for up to five years after the evaluation, after which it will be deleted.

Who will the data be shared with?

The data will only be used by the evaluation team, unless you tell us something that may harm you or others, in which case we may have to inform a person of authority. As a GPE evaluation, GPE have a right to the data from the evaluation, but these data will be anonymized.

Do I have to do this interview?

No, it is a voluntary interview. You are free to refuse to answer any question or to withdraw from the discussion at any time without repercussions. You may also request that your data be removed from the evaluation after the interview.

After reading through this document, your consent to participate an interview will be confirmed if you agree to take part and confirm your availability with the data collection team.

Your interviewer will also confirm in the interview that you agree to take part and make reference to this document, and allow you to ask questions, before commencing the interview.

If you have any questions about this process or about what we will do with this data, please get in touch with the evaluation team, through the Evaluation Manager at: jessica.chu@tripleline.com

Annex 9. Quality Assurance Framework

Introduction

This annex describes the framework for how QA will be managed as part of the Country-Level and Thematic Evaluation. Ensuring quality and rigor throughout this evaluation is critical to guarantee not only the validity of its findings, but to support its credibility and therefore usefulness. Key evaluation stakeholders should be confident in the findings and recommendations put forward by this evaluation in order to be enabled to learn and make key decisions.

As part of the inception phase, we identified a number of quality challenges (see table below), which are inherent to conducting a large-scale, longitudinal and complex evaluation, as well as quality expectations set out for the evaluation.

We therefore designed a bespoke framework to respond to the quality challenges identified. The QA framework includes descriptions of the principles and standards for quality in the evaluation, roles and responsibilities for QA, the mechanisms used (including processes, procedures and techniques) and finally a QA checklist which will be used to support QA.

Quality Challenges

In the inception phase, we identified the following challenges that might affect the quality of the evaluation.

Quality challenge	Description of challenge
Quality of methods and tools	
Fidelity of evaluation approach as applied to country-level case studies	The ability to conduct cross-case analysis for our country-level case studies will require a certain level of fidelity to the overall research approach. This will be a challenge given the number of country-level case study teams conducting the research. While there will be overlap in team members (meaning there will not be 15 distinct and discrete teams), there will necessarily be a wide range of team members in order to ensure that we are drawing upon country-specific experience and expertise and to ensure that studies can be conducted simultaneously given the planned timeline for each phase.
Design of context-specific country-level case studies, including context-appropriate data collection tools	Although there will be a need to ensure consistency in the way in which the overall evaluation approach is adapted to the country-level, there will also be a need to ensure that the study is carried out in a manner that accounts for the country context. This includes considerations such as the different country-level respondents who may be involved in the study, different languages and different cultural expectations for interview conduct. This is a separate consideration to the way in which the study will need to be adapted, in terms of its objectives and key evaluation questions, to account for the country-specific theory of change.
	In particular, stakeholder engagement processes and interview guides for key-informant interviews will need to be tailored, but still validated for their effectiveness both for use in-country and for achieving the same overall research objectives.
Quality of data and evidence	
Ensuring the safety of respondents and researchers across a large number of country-level case studies	While country-level case studies will require adaptation with regard to fieldwork plans and logistics, all studies are subject to the same standards and principles with regard to ethical conduct and preventing harm. This expectation (set out in the research guidelines and Triple Line policies) is critical to ensure that the safety of evaluation participants (e.g., respondents) and researchers. This is required in

	order to meet our ethical research standards as well as to support the collection of valid data and evidence.
	This will require detailed procedures for risk assessment and recruitment. In addition, it will require that all evaluation team members (particularly country-level case study team members) will be required to undergo a standardized training process on safeguarding and ethics in order to adhere to the policies and expectations set out for conduct.
Ensuring the validity and precision of the data collected	For our findings to be valid, we need to ensure not only that the right data are collected from the most relevant stakeholders, but also that the stakeholders are able to speak freely and precisely and that findings are subsequently validated and triangulated.
	We have included a stakeholder mapping activity as part of the case study design phase, as country-level case study teams will need to determine the most relevant stakeholders for data collection (including which stakeholders might be the most knowledgeable to answer questions with precision and accuracy, and how to triangulate findings across different stakeholders).
	Country-level case study teams will also require consultation with individuals with knowledge on the political context of each case study. This includes CTLs from the GPE Secretariat, as well as the in-country researcher (part of the country case study team), who can provide guidance on how to approach respondents in order to ensure that they are clear on the intent of the evaluation so that they are likely to speak more freely.
Ensuring the timeliness of data collection and therefore of evaluation deliverables	Evaluation workplans have been designed to collect data at the "right time," in terms of accounting for the progress of roll-out and rate of change of the interventions (e.g., GPE operating model and priority reform implementation) as well as for when the evaluation deliverable is required for decision-making. However, this also produces tight timelines for this process, which can mean that there are a number of risks that, if unmanaged, could interfere with the production of timely data.
Transparency in the processes of data collection and analysis to support confidence in and credibility of findings	There needs to be transparency in data and analysis for the findings to be credible. Both the country-level and core evaluation teams need to ensure they are documenting the tools and processes used for data collection as well as the processes for analysis. These should be clearly explained and made available to evaluation users to support confidence in the findings.
Consistency in the strength of evidence and data collected across country-level case studies	In order to ensure the comparability and synthesis of findings from across country-level case studies (to support reporting on thematic case studies and the synthesis report), there needs to be consistency in the strength of the data and evidence. This includes confidence in the validity of the data and the degree to which biases and inaccuracies in the data and evidence are managed (through triangulation),
Quality of deliverables	
Consistency in depth and rigor of analysis for country-level case study reports	As with consistency of the quality of evidence and data, there needs to be consistency in the analysis (including depth and rigor) in order to produce credible and comparable findings. This can be a challenge when different teams or individuals are involved in the analysis process.
Consistency in presentation for reports and analytical outputs	To deliver this longitudinal and complex evaluation, we will require a large team who brings to the evaluation different expertise and knowledge. This poses some difficulty in ensuring that the evaluation

> and its outputs are delivered in a consistent manner. This is particularly important given the large number of country-level case study reports.
>
> Evaluation deliverables and outputs (including analytical outputs) need to be consistent in their presentation in order to be usable.

Key Quality Principles

We have reflected on the quality challenges identified above in order to define a set of principles that we will use to support quality throughout the evaluation:

- **Co-design and collaborative approach:** to support quality in methods and outputs in terms of credibility and validity, we will use a co-design and collaborative approach. This includes allowing for time in our workplan to ensure that we can collaborate with both the GPE Secretariat (particularly the R&P team) as well as with our evaluation participants to ensure that the evaluation design (including at the country-level) is appropriate and that data, evidence and emerging findings are valid.

- **Proportionality:** the evaluation is ambitious in its scale and objectives and seeks to be consultative. Therefore, at times throughout the evaluation, we may uncover additional learning needs or interests. While we will endeavor to incorporate these, we can also run the risk of stretching the evaluation's resources and attention and diluting the effectiveness of the evaluation. Therefore, to ensure the quality of evidence and deliverables in terms of consistency, we will apply proportionality to ensure that we are prioritizing investments (in time and resources) to address the most critical questions for the achievement of the objectives of the evaluation (as outlined in the evaluation matrix). We will also use this principle to make decisions on methodological approaches to ensure that we are collecting a proportional amount of data and dedicating an appropriate amount of time for analysis.

- **Flexibility:** given the longitudinal nature of the evaluation and its large geographic scope, we need to be flexible to ensure that our evaluation can adapt to any changes in contexts or as interventions evolve (e.g., GPE's operating model or country-level priority reforms). This will allow us to ensure the quality of methods in terms of appropriateness.

- **High standards of ethics and safeguarding**: to ensure quality in data in terms of accuracy and integrity, we need to ensure that data are collected with the highest standards of ethics and safeguarding in order to protect both evaluation participants and researchers. Evaluation participants should be free to participate without fear of safety or reprisal.

- **Usability:** to ensure that our evaluation outputs are usable by key evaluation stakeholders, there needs to be quality in our deliverables in terms of being concise and engaging. We will ensure that all reports are clear, simple, free from jargon and short. They will be focused on communicating key findings and takeaways, tailored to their targeted audience. We will also produce additional outputs (such as annexes to reports) to support the credibility of our findings by ensuring that there is transparency in our research and analytical methods – these will also be made publicly available.

Quality Assurance Approach

With the quality challenges and principles identified above, we devised a QA approach that will support us to manage quality throughout the evaluation. We built QA activities into the workplans for each evaluation deliverable as well as throughout the evaluation process.

Roles and Responsibilities for Quality Assurance

Triple Line, as the contract holder, is responsible for upholding this QA approach. Therefore, the Evaluation Manager and Project Director play a critical role in managing the QA framework.

Role	Responsibility
Evaluation Manager	Responsible for the design and implementation of the QA framework, including for all evaluation team members to be trained on and comply with the framework
	Responsible for monitoring the quality register
	Responsible for reviewing the effectiveness of the QA framework and where necessary updating it to ensure its continual relevance and appropriateness
	Responsible for using the QA checklist when applied to synthesis activities
Project Director	Ultimate responsibility for QA in the evaluation. Responsible for ensuring that the Evaluation Manager is supported to implement the QA framework
	Responsible for approving changes to the QA framework
	Contributes to QA as a Quality Assurer and for final approval of all contracted deliverables
Core management team members	Responsible for reviewing QA methods and ensuring the appropriateness for the evaluation, with a particular focus on respective areas (suitability for technical subject areas, evaluation design and methods, and learning activities)
	Contribute to QA as Quality Assurers and support sign-off on evaluation deliverables
Country-Level Case Study Lead	Responsible for ensuring that QA methods are appropriate and are integrated into the country-level case study methodology
Country-Level Case Study Manager	Responsible for ensuring that QA methods are integrated into the country-level case study workplans
	Responsible for monitoring the use of the QA checklist throughout the process of country-level case studies
Case Study Lead(s)	Responsible for implementing the QA processes in country-level case studies and using the QA checklist

Mechanisms Used to Support Quality Assurance

We will use several mechanisms to support QA throughout the evaluation, including the following:

- **Use of templates:** throughout the evaluation, we will create templates that will act as guides to support the development of evaluation outputs. This includes for any desk research or analytical outputs (such as the PEA and stakeholder map) for country-level case study design (such as theory of change and assumptions and evaluation matrix) for data collection tools and plans (including interview guides, fieldwork plan and stakeholder consultation strategy), for data entry (survey and note-taking templates for qualitative interviews), for slide decks for presentations and workshops, and for reports and annexes. Although teams will be allowed to adapt some of these templates as necessary, the templates will ensure that the correct quality expectations are set, and to support consistency.

- **Use of workshops and verbal feedback:** our approach is collaborative, recognizing that each team member brings expertise to the assignment, and that we have much to learn from colleagues. We have built in opportunities throughout our evaluation for peer review and challenge (both across country-level teams, and from our core management and evaluation teams), in order to collaborate on and exchange ideas, which will support consistency and

rigor. This includes the country-level mini inception workshop and the comparative case study workshop.

- **"Fresh pair of eyes" reviews:** these are informal review processes used within country-level case study teams and in the core evaluation team to ensure that all evaluation outputs (including tools, plans and analytical outputs) are reviewed with a "fresh pair of eyes." This is appropriate for reviews of draft materials and outputs as well as for non-deliverables. These are likely to be done within the core evaluation team, as well as by core evaluation team members to support to country-level case studies to help bring in consistency and learning across studies and to the thematic case studies and synthesis process.

- **Technical or thematic expert reviews:** these ensure that our evaluation outputs are of a high standard with regard to technical content. We will use key expertise across the team for these reviews or, where there are gaps in our team, draw in support from external thematic advisors. For example, our thematic case studies will undergo additional review processes by gender and domestic financing experts to ensure the validity of our findings.

- **QA reviews:** these are formal review processes used primarily by the core management team to provide sign-off on key evaluation deliverables (that is, those listed in the evaluation's scope of work). All key evaluation deliverables will require sign-off from at least two Quality Assurers, which will be allocated across the core management team. This will help reduce the QA burden on one individual and help to contribute to consistency of quality standards by ensuring that Quality Assurers are able to consider and compare quality across multiple deliverables. The Project Director will have final sign-off on key deliverables, such as the synthesis reports and synthesis findings slide decks. Time and resources have been built into the workplan and budget for these review and sign-off tasks.

- **External reviews:** working collaboratively with R&P, we will solicit reviews beyond the process for client sign-off of deliverables. These are proactive efforts to collaborate with R&P to ensure that, where we are unsure, evaluation outputs are relevant to the context of GPE.

- **External consultations and validation workshops:** we have also incorporated into our methodology the use of external consultations and workshops to support the validation of our evaluation efforts. We will use these both to check the appropriateness of our country-level case study design (through initial consultations with CTLs) as well as to validate our findings at the country-level (country-level data collection debrief workshops) and at the synthesis level (emerging findings presentations and recommendations co-creation workshop).

- **In-house or professional editing services:** to support the quality of our deliverables, we will draw in support from either in-house or external professionals, as relevant. Our in-house team can provide editorial and graphical support, as well as editing and proofreading. For key deliverables, we will draw upon the services of our partner, Scriptoria, an award-winning communications company that specializes in communications in international development. Scriptoria will provide professional editing and publishing services as well as professional translation. In both cases, all editing will be done using GPE's editorial style guide.

Quality Assurance Checklist

The QA checklist below will be used throughout the evaluation in order to set and manage quality expectations. All team members will be socialized on the checklist, which, for country-level case studies, will be completed by the country-level case study team lead in an ongoing process throughout the case studies. The Country-Level Case Study Manager and Evaluation Manager are responsible for monitoring the use of the checklist throughout the country-level case studies. At the synthesis and learning-level, the Evaluation Manager is responsible for completing the checklist, with support from the Project Director to ensure compliance with this practice.

Category	Quality criteria
Quality of methods and tools	1. Country-level case studies have been appropriately tailored and adapted to country contexts, and include the following: • A strong understanding of the local context, demonstrated through the completeness of the political economy analysis • Tailoring to reflect the content and focus of the country's proposed transformational reform, demonstrated through the validity and completeness of the country-level theory of change, assumptions and evaluation matrix • A consideration for the sociopolitical dynamics within the country, including considerations for gender, as demonstrated through the completeness of the stakeholder map and the inclusiveness of the data collection protocols.
	2. Country-level case studies are designed with fidelity to the overall evaluation approach and to generate evidence and results that contribute to the overarching evaluation matrix.
	3. Country-level case studies have identified the appropriate participants as well as the appropriate channels for reaching respondents as demonstrated in the stakeholder map and stakeholder consultation strategy.
	4. Country-level case studies include a diversity of participants, including where possible and relevant, proposed respondents who represent or can speak on behalf of end users, marginalized and vulnerable groups (women and disability etc.), as demonstrated in the stakeholder map and stakeholder consultation strategy.
	5. Country-level case study data collection strategies and protocols are designed to be culturally and context-appropriate, including translation where necessary.
	6. Country-level case study data collection strategies and tools are designed to be proportional, practical and feasible, where proportionality concerns the amount of data collected (e.g., to answer the respective data collection) and the time asked of respondents. Interview protocols contain questions that are phrased in a way that are open-ended, easy to understand, do not cause bias and are an appropriate length.
	7. Derivations from the overall case study approach or templates are clearly explained.
Quality of data	8. Data collection has been conducted in a way that is free from conflict of interest and outside influence. Any limitations to the quality of data collected, or any threats to the independence or impartiality are initially discussed with R&P and then clearly explained to the evaluation audience in the respective report.
	9. Data collection has been conducted in a way that complies with the research protocols for this evaluation, including in an ethical manner, is gender-sensitive and human rights-responsive and with considerations for safeguarding.
	10. Methods used for data collection and the results of data collection were made transparent and clearly explained to the evaluation audience in the respective report. This includes an explanation of the sampling frame (what data were collected from whom and when) and how data were collected (with what tools and with the justification for the selection of tools). Any changes to the methodology or fieldwork plan are clearly explained and justified.

|		
	11.	The quality of data is assessed against the suitability to serve the intended purposes, including validity/accuracy, reliability, precision, timeliness, integrity and completeness.
		Where there is uncertainty against any of the above criteria that may affect the validity of the findings, this will be initially discussed with R&P to determine an amenable resolution and where relevant, explained in the corresponding report.
	12.	All data collected are de-identified and stored securely, including password protection where necessary.
		Any keys used to link data to respondents are stored separately and password protected.
	13.	All data collected are organized and stored in such a way that they can be recalled for re-analysis.
	14.	(Secondary) quantitative datasets have been cleaned and checked for completeness, duplication, errors or inaccuracies and unexpected responses.
Quality of evidence	15.	Data have been analyzed in a way that is inclusive. For example, data analysis considers the perspectives of different stakeholders.
		Data have been analyzed in a manner that is gender-sensitive and considers other forms of marginalization. Data are disaggregated where possible.
	16.	Data have been analyzed in a structured and systematic way.
		For example, data analysis follows logically from the evaluation matrix and systematically against the evaluation questions. The findings clearly and transparently link to conclusions (and recommendations where relevant).
	17.	Findings are triangulated using multiple (and where possible mixed) sources of data and realist evaluation principles.
		As a result, findings consider a variety of perspectives, present concurring patterns as well as contrasted events, clearly and adequately explain why/why not and how certain phenomena were observed or not.
	18.	Unintended or unexpected findings (positive or negative) are presented and explained.
	19.	Gaps in the analysis and in the strength of the evidence are reported and discussed.
	20.	Conflicting findings or divergent perspectives are presented, with explanations for these where possible.
	21.	Findings have been validated by a range of key stakeholders.
	22.	The report includes conclusions that present reasonable judgments and are substantiated by the findings. Conclusions reflect upon the evaluation questions and judgment criteria set out in the evaluation matrix. The report presents clear linkages between the findings and conclusions.
Quality of deliverables	23.	Evaluation outputs are written clearly and concisely with an impartial and constructive tone.
	24.	Evaluation outputs follow the specified template, which provides a coherent and complete structure for the report. This includes a clear presentation of an introduction and purpose/objectives, any background or context, before presenting content. Output content is followed by conclusions or recommendations where appropriate. Finally, outputs will also include short (2–3 pages), standalone executive summaries and have been translated into French.

Annex 10. Team Organogram and Roles

This annex provides some elaboration on the evaluation team structure and named members, as at the time of the inception report. As country-level case studies are launched, the overall evaluation team will evolve and grow as more team members will be added to support these studies. Therefore, the structure of the team (depicted below) is designed to support the growth of the country teams by providing distinct roles within the core evaluation team to support country-level case study teams, as well as having overlaps in members between the core evaluation team and country-level case study teams, as well as across different country-level teams.

In addition, the team will draw upon Triple Line's internal operations team for support to operationalize the country-level case studies (such as duty of care and logistics support), as well as additional thematic expertise as required.

Responsibility for personnel decisions for the evaluation team are made by the core management team. If and when changes are made to the team (such as to replace team members or as the team grows), these changes will be shared with R&P through weekly check-ins (including the sharing of CVs where relevant).

Structure of evaluation team
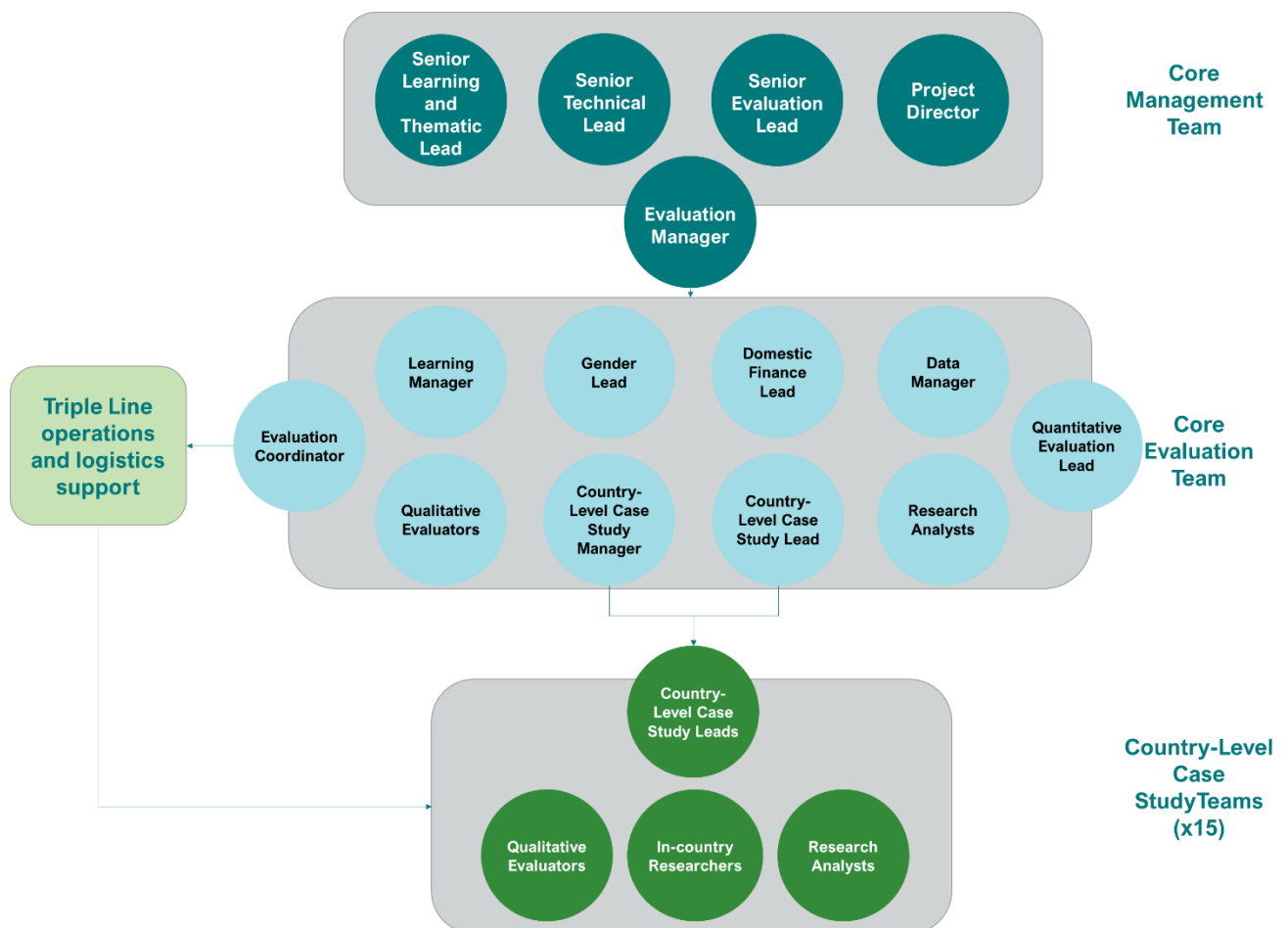
Personnel of the core management and evaluation teams

Team	Name	Role(s)	Organization
Core management team	Rebecca Allinson	Senior Evaluation Lead; Quality Assurer	Technopolis
	Jessica Chu	Evaluation Manager	Triple Line
	Michelle Kaffenberger	Senior Technical Lead; Quality Assurer	Independent
	Lorenzo Newman	Senior Learning and Thematic Lead; Quality Assurer	Learn More
	Clarissa Poulson	Project Director; Quality Assurer	Triple Line
Core evaluation team	Peeyush Chaturvedi	Evaluation Coordinator	Triple Line
	James Handley	Data Manager; Qualitative Evaluator	Triple Line
	Adam Krcal	Country-Level Case Study Lead; Qualitative Evaluator	Technopolis
	Mahima Mehra	Research Analyst	Triple Line
	Giorgio Monti	Quantitative Evaluation Lead	Learn More
	Alice Pelosi	Gender Equality Thematic Lead; Qualitative Evaluator	Learn More
	Mamta Raichura	Research Analyst	Triple Line
	Elizaveta Rusakova	Qualitative Evaluator	Learn More
	Ekaterina Shaleva	Country-Level Case Study Manager; Qualitative Evaluator	Triple Line
	Giovanni Zino	Learning Manager; Qualitative Evaluator	Learn More
	TBD	Domestic Financing Thematic Lead	TBD