*==DRAFT FOR DISCUSSION PURPOSES ONLY – NOT FOR CITATION==*

# THE LEARNING DATA COMPACT
# CONCEPT NOTE[1]

# Table of Contents

# Abbreviations and acronyms

EAGLE      Education Analysis for Global Learning and Equity
EMIS      Education Management Information System
ERCE      Regional Comparative and Explanatory Study (Estudio Regional Comparativo y Explicativo)
GAML      Global Alliance to Monitor Learning
GEPD      Global Education Policy Dashboard
GPF      Global Proficiency Framework
ICAN      International Common Assessment of Numeracy
IEA      International Association for the Evaluation of Educational Achievement
IMG      Independent Monitoring Group
INDS      Integrated National Data System
LDC      Learning Data Compact
LSMS      Living Standards Measurement Study
MDG      Millennium Development Goals
MICS      Multiple Indicator Cluster Surveys
MICS FLM      MICS Foundational Learning Module
MILO      Monitoring Impacts on Learning Outcomes
MPL      Minimum Proficiency Level
OECD      Organization for Economic Co-operation and Development Programme for the Analysis of Education Systems
PASEC      (Programme d'Analyse des Systèmes Éducatifs de la Confemen)
PIRLS      Progress in International Reading Literacy Study
PISA      Programme for International Student Assessment
SDG      Sustainable Development Goal
SDI      Service Delivery Indicators
SEA-PLM      Southeast Asia Primary Learning Metrics
TCG      Technical Cooperation Group
TIMSS      Trends in International Mathematics and Science Study
UIS      UNESCO Institute for Statistics

# Technical Glossary

| | |
|---|---|
| Global Proficiency Framework | The Global Proficiency Framework (GPF) includes learning standards in reading and mathematics to define minimum proficiency levels that students are expected to have achieved at the end of each school grade level. The framework provides a common reference point to enable countries to use the results of national, regional, and international assessments to report progress on SDG 4.1.1. |
| Policy Linking | Policy linking (PL) is a standard-setting methodology for learning assessments that allows expressing student learning outcomes from existing assessments to the proficiency levels described in the Global Proficiency Framework and report progress on SDG 4.1.1 (Proportion of children and young people (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex). The methodology involves workshops with teachers and curriculum experts to determine the content alignment of learning assessments to the learning standards described under the Global Proficiency Framework and, in turn, express student learning outcomes in the proficiency levels of the Global Proficiency Framework. |
| Student linking | Student linking (SL) follows the psychometric tradition of statistically linking scores from different assessments by administering the two assessments to a common sample of students. This methodology permits to express scores from different assessments in a common scale of student achievement, facilitate comparison of scores, and interpretation of results. One use of this score linking design is the benchmarking of results from national assessments to those from regional and international assessments. |
| Global Item Bank | The Global Item Bank is a public repository of test items developed by the UIS and the World Bank to help countries generate learning data on reading and math. |

# 1. Objective

*The Learning Data Compact aims to improve the availability, frequency, relevance, and timeliness of learning data. Its main objective is to radically expand the access to and use of learning data for policy makers, educators, and other stakeholders in low- and middle-income countries. This initiative likewise aims to help establish sustainable capacity to effectively produce and use such data for informed decision making to improve teaching and student learning outcomes, and help countries achieve their educational goals.*

**Gaps in Learning Data are a Systemic Problem**

Learning data is vital to inform effective education policy and practice. These data insights are key to understand children's learning levels and how they are learning, which can inform *action* to improve learning. But efforts to collect learning data are fragmented and irregular, especially in low- and middle-income countries. For example, in some countries, learning assessments have only been carried out once, often with no temporal comparability or clear pedagogical interpretation for policy insights. Even when such data are collected, they may be out of date by the time they are made public, as it can take two to three years to report the results. Frequently, the data quality varies, data comparability is limited, the microdata is poorly documented, and not easily available for supplemental research and analysis.

While there is urgent need for learning data, particularly in the wake of COVID-19, there are several reasons why comparable data are not produced and used frequently. Partly, low quality and limited use of existing data, and inadequate financial and technical capacity, drives low demand from policy makers, educators, and other stakeholders. Political economy constraints may further limit data collection, and the publication and release of data. Furthermore, lack of frameworks to describe the necessary priority data, each assessment's purpose, and tools and frequency with which data should be collected, may further inhibit its production and use. Often after a financially and time-consuming data collection process, the data's benefits may appear low because only a single "official" report is produced, and not much is done to promote and enable data reuse. Furthermore, there may be little integration of different types of data, especially to link large-scale learning assessment data with formative and classroom assessments and with other statistical activities, such as school surveys and administrative records in an education management information system.

**Goals of the Learning Data Compact**

With a focus on low- and lower middle-income countries, the Learning Data Compact (LDC) aims to boost processes for regular and timely learning data by offering a menu of evidence-based methodologies, tools, and solutions. Developing countries can deploy flexible strategies to improve the quality, relevance, timeliness, and cost-effectiveness of information on learning to inform decision-making at the teacher, school, subnational, and national levels.

The LDC provides a framework for collaborating processes, initiatives, and resources to measure learning, and increases their joint impact by avoiding fragmentation and duplication of effort and resources. An aligned effort to measure learning can improve the data's relevance to country contexts, maximize technical and financial resources to ensure transparent and accountable processes in data collection, and enhance country benefits and demand for data through its uses. It also helps mobilize future resources and align regional and global public goods based on country needs. The World Bank, UNESCO, and UNICEF have collaborated under shared principles, shared processes, and shared accountability (shown in Figure 1 below, and explained in following sections), to push forward a learning ecosystem that allows alignment across development partner and country objectives.

**Figure 1. Shared Principles, Processes, and Accountability Under the LDC**
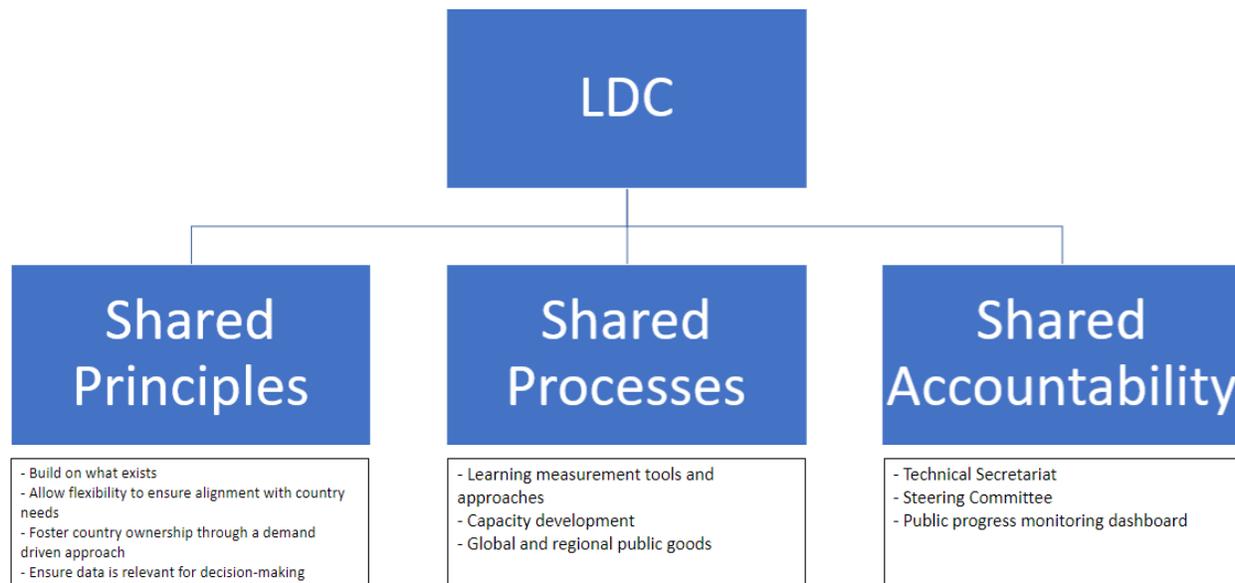


For participating countries, the LDC aims to:

- Support greater alignment between programs and partners on how to plan, design, implement, analyze, and use results of large-scale learning assessments, building on existing solutions and the latest innovations.
- Promote measures of student learning that can be repeated over time, and when relevant compared across countries.
- Strengthen the link and integration between large-scale student assessments, data on drivers of learning (including through household surveys, school surveys), and classroom assessments.
- Improve production and use of data to inform action on the ground by strengthening existing institutional capacities.

As a global initiative, the LDC aims to:

- Strengthen quality, oversight, transparency, and impact of global efforts to measure student learning outcomes.
- Strengthen coordination among different partners to support country systems in measuring and monitoring learning.
- Complement regional and international assessment initiatives including PASEC, SEAMEO, SACMEQ, PILNA, LLECE, PIRLS, TIMSS, and PISA.

See Figures A1 and A2 in the annex for the theory of change for the LDC at a national and global level respectively.

To achieve these aims, the LDC proposes to support, in a coherent and integrated manner, the use and collection of data for three key purposes:

1) **Measure learning progress (Pillar 1):** This can be done by conducting at least two rounds, of temporally comparable large-scale learning assessments in at least two grades and subjects

(2x2x2) by 2030. For these assessments to inform teaching practice, they should have pedagogical interpretability and relevance for teachers in the country, and be linked to standard-setting processes (identifying scores against proficiency levels) in the country.

2) **Measure the drivers of learning (Pillar 2):** This can be done by integrating learning modules for different age groups (children, pre-school, early-grade, end-of-primary, lower-secondary, and adult) to traditional multitopic household surveys, such as MICS and LSMS; promoting the integration of administrative records to understand the drivers of learning; and deploying school surveys, such as the Global Education Dashboard.

3) **Understand classroom practices (Pillar 3):** This can be done using classroom assessments to inform diagnostics of learning and inform teaching practices. Such assessments can be used by teachers to implement, for example, Teaching at the Right Level techniques which ensure that instruction is targeted to the level of the child. Such assessments can also enable teachers to understand progress against curricular standards and goals.

Cross cutting these three types of data is the commitment to strongly support data use and re-use (Section 4 describes the 3 pillars in greater detail). That requires building in-country data analytical capacity and fostering data integration in the ministries of education and other local agencies and doing this safely and transparently[1]. The purpose of this concept note is to detail how we can collectively address the need for a better learning measurement ecosystem, this is very much a living document, and we expect it to evolve as our partnership expands and matures. The following sections proceed as follows: Section 2 describes the LDC's broader vision and rationale in building on the Sustainable Development Goals 4.1. to strengthen country systems. Sections 3, 4, and 6 discuss the shared frameworks between collaborating institutions under the LDC, namely of principles, processes, and accountability. Section 5 simulates costs to attain measurement goals for the three pillars, and Section 7 expands on the LDC's governance and operational model.

## 2. The Vision

***Why the LDC, and Why Now?***

In 2015, countries committed to achieving Sustainable Development Goal 4.1: ensure that by 2030 all girls and boys complete free, equitable, and quality primary and secondary education leading to effective learning outcomes. Previously, the Millennium Development Goals (MDGs) had only indirectly focused on quality, such as by centering on raising completion and reducing the number of children out of school. The SDGs focus on learning and emphasize learning quality as well as learning inequality. Consistent with SDG 4.1, the LDC continues to monitor these learning points for SDG reporting, but a better measurement ecosystem is required to allow global and comprehensive participation. The ongoing pandemic and joint institutional efforts to understand its effects has created a window of opportunity for the LDC to be implemented.

The extent of the learning crisis is well understood. Globally, 53 percent children are learning poor—they cannot read and understand a simple age-appropriate text at age 10 —and this number reaches 90 percent in Sub-Saharan Africa. The Learning Poverty indicator, created by the World Bank and UNESCO Institute for Statistics[2], captures educational access and quality by combining schooling and learning in a single measure. Analysis based on the indicator suggests that most Learning Poverty is driven by school-age-children who are attending school but fail to reach the absolute minimum level of learning to be productive citizens in the economic, social, political, and cultural spheres that their countries aspire to. Moreover, although learning levels have improved over the past two decades, progress has been slow and uneven across countries. Even prior to COVID-19, projections in a "business as usual" scenario

suggested that Learning Poverty, can fall 10 percentage points by 2030, reaching 43 percent, with the world reducing Learning Poverty by less than one percentage point on average.[3]

The sheer inequality in learning is less understood. Although the learning gap between poorer and richer countries is often talked about, the reality is that about half the global learning inequality is within countries, and in many countries, it is within the school and even classroom. Consider this: as high as 80% of 10 year olds (or students at the end-of-primary, as captured by the SDG 4.1.1b indicator) in poor countries, cannot understand a simple age-appropriate text, implying they may not be able to read the most common words in their home language. It is hard to imagine how any of this policy challenges will be addressed without a solid measurement system, which enable an accurate understanding the magnitude of the crisis, its nature and context, and implications in the classroom practices.

COVID-19 has exposed cracks in our education systems. These inefficiencies affect children at various education levels but can be particularly consequential when they show up at foundational skill levels. Children who are not already instinctual readers or who do not have a good intuitive, conceptual, and procedural grasp of simple mathematics, are likely to be suffering acute learning losses. Sadly, in many low- and lower middle-income countries, pre-COVID learning measures often cannot be interpreted in clear pedagogical terms. Though the pandemic has highlighted the fragility of learning systems, it has also given a new impetus to the learning measurement agenda. The expected learning losses, with estimates largely from simulations so far[4], need to be measured to help countries allocate the resources and effort to address the learning crisis and identify mitigation strategies that are likely to be successful—and that could be replicated, improved and scaled up for a stronger learning recovery and acceleration. As schools reopen, it is also critical to enable teachers to have the classroom resources to rapidly identify children's learning levels and potential learning gaps in order to help students catch up.

It is possible to move the needle on learning – and do so quickly through solid partnerships around the production and use of data. Recent game changers offer a new window of opportunity for rapid progress on the learning measurement agenda (see Figure 2):

- Since 2019, countries and development partners have finalized and endorsed a Global Proficiency Framework (GPF) that defines the SDG-aligned minimum proficiency levels.
- The GPF agreed that countries can report their learning progress flexibly, using the content of different learning assessments implemented in the country or regional levels, which can be aligned with minimum proficiency thresholds using methodological innovations, such as policy linking or student linking.[5]
- New tools were developed for the GPF, such as the South Asian Learning Primary Metrics, the extended UNICEF foundational learning module, and the OECD's new short PISA module.
- Many regional assessment programs, such as ERCE and PASEC in Latin America and the Caribbean and Sub-Saharan Africa, are reaching standard protocols and consistent country participation, allowing temporal comparison of their results.
- Last, COVID-19 has led to additional innovations, such as the Monitoring Impacts on Learning Outcomes (MILO) program piloting new assessments that build on the GPF and can rapidly report results on a country's performance or learning losses with respect to the GPF.

Furthermore, many donors and development partners worked closer together in response to COVID-19. For example, UNESCO, UNICEF, and the World Bank partnered to successfully co-lead several efforts and initiatives focused on COVID-19 response in addition to the LDC, such as the Joint Survey of Education Ministries and Mission Education Recovery.

**Figure 2. The Learning Data Compact: Why now?**

Game changer developments since 2015
(which started to yield results in 2019)

New tools to measure learning

Better alignment

Covid-19 /learning losses

Standards/Definitions (GPF/Bank of Items)

Country-level demand

Methodologies

Coordination
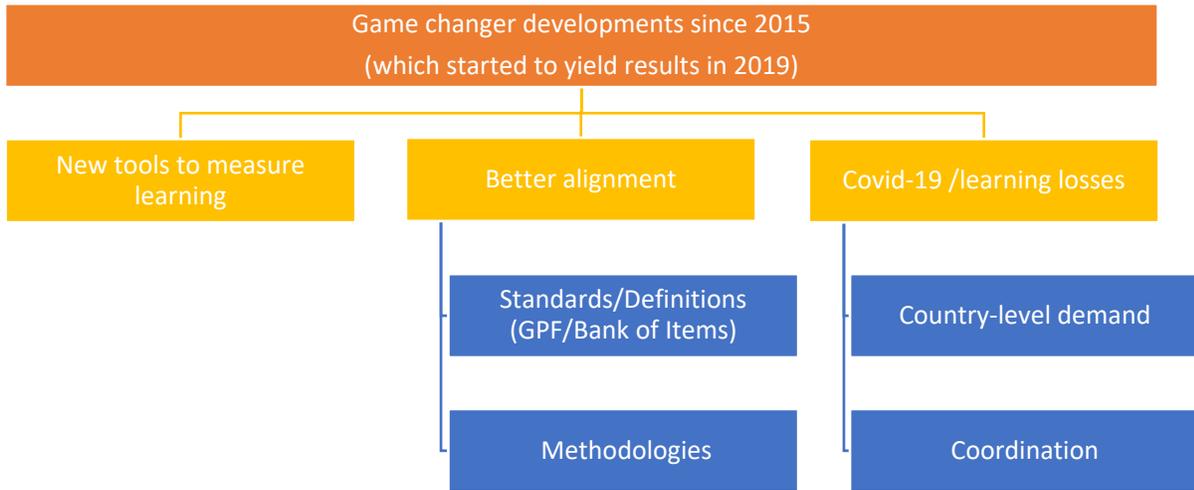
*Enter the LDC: A Holistic and Programmatic Approach to Measure Learning*

The Learning Data Compact (LDC) aims to boost the availability of regular and timely data by offering a menu of evidence-based methodologies, tools, and solutions that developing countries can deploy in a flexible manner to improve the quality, relevance, and timeliness of information from learning assessments. The LDC's approach is holistic because it will encompass not only learning data produced in large-scale learning assessments but also complementary data to understand the drivers of learning and classroom assessment data to transform the students' learning experience. And it is programmatic because change in learning outcomes is seen as being best achieved through interconnected projects with common objectives. The LDC aims to foster a measurement ecosystem that encourages collection and use of data to inform effective decision-making (see Figure 3).

**Figure 3. The Measurement Ecosystem**

Measuring Learning in the Classroom (Pillar 3)

Tracking Education System Progress (Pillar 1)

Understanding Drivers of Learning (Pillar 2)

Better Policies and Practices in School
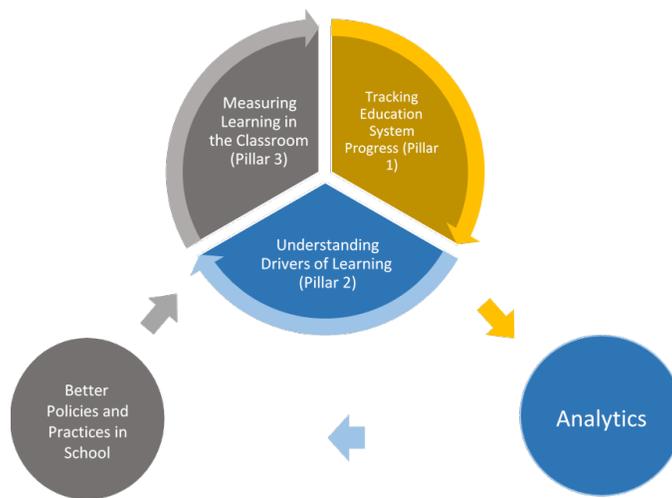
Analytics

**Data Collection**

A critical element of the LDC's vision is recognizing that the data production cycle is an integral part of the measurement ecosystem, where data analytics and designing better practices at the school are complementary components. To that end, it is critical to collect different types of learning data through classroom, household, and large-scale assessments, which can inform what and how children are learning at home and through their education systems. For example, measurements at a classroom level can look like monthly simple "mastery" checks to identify children that are falling behind. Classroom measurements also create opportunity to measure higher-order thinking skills that a standardized assessment cannot measure. Teachers can collect such data with guidance and support from coaches or visiting supervisors. Similarly, household assessments can provide useful information about drivers of learning, while large-scale assessments are useful for tracking system-level progress.

Furthermore, it is not enough for a country to implement a learning assessment as a one-off activity, which does not generate sufficient data to monitor progress and therefore really understand the drivers of learning. One-off data collection strategies might also create insufficient demand to justify the training and eventual retention of professionals who now have the skills to produce and use learning data. The measurement ecosystem encourages countries to plan and budget for multiple rounds of data collection with sufficient regularity that can justify the training and retention of staff with appropriate skills, build in-country capacity for analysis and communicate results, and ensure that policy makers habitually use learning data to inform critical policy decisions. It also allows countries to compare changes in learning over time.

**Data Analytics**

Measuring will not by itself correct poor learning outcomes, no matter how many times we measure. The data production cycle should not only have formal measurements *of* learning, such as assessments, but also *for* learning (using data to get information that can support better decision-making and teaching in classrooms). For that, different data sources need to be linked in a meaningful way. It is important to couple the data from different types of assessments to **data on context and practices,** especially to inform insights on drivers of learning. For instance, data from global or national large-scale assessments should be conceptually and pedagogically linked to national classroom assessments that are closer to how teachers teach in the classroom, which should be linked to curricular standards and goals.

Such **formative tools** should be part of a pedagogically and administratively coherent package. Particularly useful are coach or deputy principal data (or data supplied to them) on teachers' classroom practices, the support offered to teachers, whether teachers find that support useful, and whether the support makes a difference in teaching practices. These factors are linked to practical contexts such as how teachers are teaching, which also implies evaluating how systems are managed. Are curricula and lesson plans designed to support improve learning? Do certain assessments align with curriculum standards and goals? Do teacher practices and lesson plans align with curriculum and assessments? For example, Nepal's national learning assessment includes contextual questionnaires to students, teachers, and headteachers across 1400 schools. Lower assessment scores and questionnaire findings suggested a misalignment between the intended national curriculum and how they were delivered. As a result of a combined measurement strategy, reports were able to recommend better teaching methods and learning environments across diverse demographic needs.[6]

**Data Use for Better Policies and Practice**

The measurement ecosystem allows collection of different types of data that can inform analytics to inform better decision-making by teachers, managers at different levels (school, subnational, national),

and policymakers. The ecosystem invests proportionally more in the use and re-use of data than on its production to improve teaching practices and policies. See Box 1 for some principles underpinning this approach. A well-integrated measurement ecosystem can help schools with self-evaluation and review, for example, by providing policymakers, teachers, and parents, with detailed information on student progress in relation to curricular standards, and help provide support as needed. Learning assessments that are good for global or even national reporting are simply not sufficient on their own to understand required changes in policies or practices and support teacher actions in the classroom, which are essential for progress. Similarly, teachers also need the larger assessment system to understand the broader systemic issues. Additionally, it is important to train teachers – through both pre-service and in-service professional development - on the basics of how to develop and implement assessments and critically, how to use assessment data to inform their teaching. Similarly, principals and school leaders need to be trained in how to support teachers on developing and implementing assessments and use the information to modify teaching practices.

Finally, some assessment units in countries would claim that their only duty is to provide the raw data collected and that whether and how the data get used is not their responsibility. Education authorities have a responsibility to ensure that assessment data gets used and re-used to support actual teaching practices and inform pre-service and in-service teacher professional development programs, management, as well as policies (See Box 2).

---

**Box 1. Beyond Data Collection: Facilitating the "Uptake Cycle"**

A commitment to produce high-quality statistics can improve the credibility and usability of data, reinforcing demand for such data. The "uptake cycle," the actions needed for key education stakeholders to understand and use the data, includes:

1. Review the assessment capacity and produce a strategy development tool to create an assessment profile, define the assessment cycle, and develop a plan of action.
2. Extend the dissemination strategy beyond producing a report.
3. Develop capacity in learning assessment units, national examination councils, and institutes of assessment and evaluation.
4. Develop the capacity to analyze, interpret, and use the learning assessment results to inform policy and planning, including curriculum, teaching, and learning processes.
5. Strengthen institutional capacity to link learning assessment and other processes for monitoring and evaluation coordinating learning assessment data and other sources of learning data.
6. Develop, test, and communicate effective models—such as the UNICEF Multiple Indicator Cluster Survey (MICS) EAGLE—for taking data from the system assessments to policy makers and stakeholders, and using intermediate forms of measurement such as school assessments to improve actual teaching in the classroom.

A common (but false) dilemma that many pose is "What should all the staff involved with test preparation and data analysis and reporting do between measurement cycles?" The answer is: disseminate findings, learn what happens in the uptake, work on the links to classroom-based and community-based assessments, and do research and prepare for the next measurement (Figure 4). Retaining staff for those activities is substantially easier is the country has a known schedule of data collection.

**Figure 4. The Measurement and Uptake Cycle**
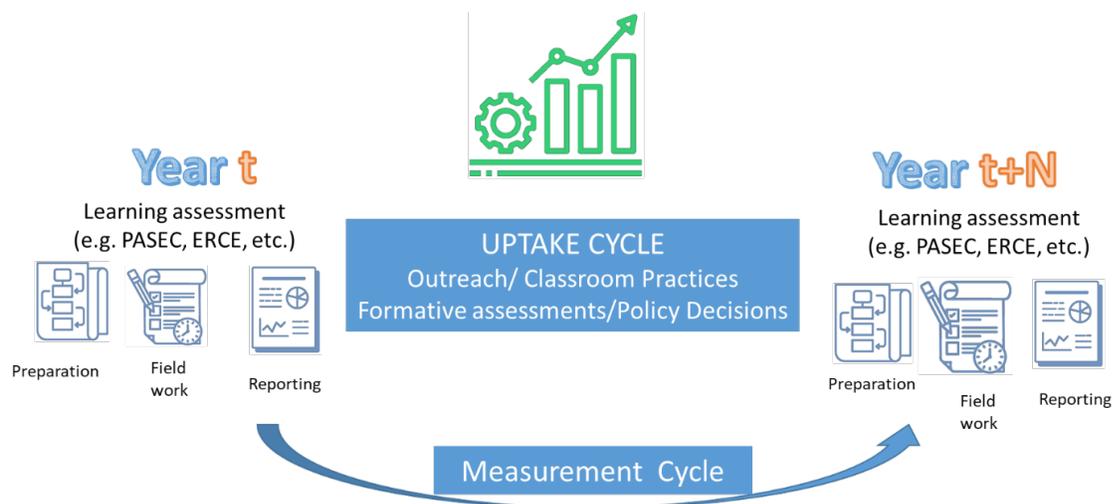


## 3. Shared Principles

To further global alignment on the LDC's measurement strategy and goals, joint institutions created three frameworks that countries, development partners, and agencies should operate under: Shared Principles, Shared Processes, and Shared Accountability. This section focuses on the shared principles to guide the choice of learning measurement, and to ensure assessment data are consistent with long-term strategic goals of effective decision-making. The following shared principles are important not just for designing assessments or deciding which assessment to buy "off the shelf," but for developing an assessment systems for one's own country. That system should be good not just for reporting but for managing improvement at all levels of education, developing the capacity to guide decision making, and linking the system-level assessments to formative assessments and classroom practices.

**Principle 1. Build on what exists**

It is key to **(develop) and build on existing capacity** of data producers, analysts and users. Most countries can build on some existing capacity. Assessing and reporting with given frequency and regularity can foster habit and expectation. Development partners can not only help fund data collection and analysis, but even more so in local capacity building, such as in the mechanics of assessment and how to make the results useful to policy makers, and especially district officials, principals, and teachers.  It is critical to ensure that activities are aligned with the national strategy and have on-going links to classroom practices, helping instill a regular habit of data generation and use.

**Principle 2. Allow flexibility to ensure alignment with country needs (not one-size-fits-all)**

It is important to know **what to assess and how to measure it**. The learning data ought to measure against a clear standard of what the learner must know, comprehend and be able to do at a specific age/grade, criteria that can be laid out in the national curriculum and/or on the Global Proficiency Framework. A good assessment system needs to align with a well-designed, coherent curriculum, and is flexible to target grades where the educational system still faces its most significant challenges. In a constrained environment, the assessment system must prioritize the assessment of segments or grades where most of their student population is, or where there are clear indicators that quality is still a significant problem.

For example, if most students do not reach lower-secondary, priority should be given to the measurement of learning at early-grade or end-of-primary. Alternatively, if a significant share of students are enrolled in lower-secondary, assessments might want to target students at the end-of-primary and lower-secondary grades. See Section 7 for further details on how countries may prioritize measurement across different grades.

In addition, the assessment should account for children's learning levels. If children cannot read the assessment instructions, the assessment needs to be designed for instructions to be read to the children, a traditional case of early grade-measures. Similarly, assessments that are too difficult may not be able to give information on the sub-skills children need to have before acquiring foundational skills.

**Principle 3: Foster country ownership through a demand-driven approach**

The approach should be **demand-driven to foster** strong country ownership. Through data reuse, it is possible to enhance stakeholders' perceived values of collecting data. Studies using existing learning data demonstrate its expansive uses to understand effective education policies. Since 2020, selected assessment programs were cited in more than 48,000 studies (see Table 1 below). For example, Mexico disseminated diagnostic feedback based on national assessments and information on school quality to help improve student learning outcomes[7]. Similarly, large-scale assessments provided information to curate user friendly reports for diagnostic feedback to schools in Argentina. The recommendations encouraged principals to use collected data for management decisions, and improved reported children engagement with teacher  instructional activities.[8] It is particularly clear that some assessment programs are significantly underutilized, especially in regions such as Sub-Saharan Africa.

**Table 1. Number of Google Scholar Citations per selected International, Regional or National Learning Assessment Programs**

| | Assessment Program | First year | latest Round (available) | Number of Rounds | Latest Number of countries | Data Access[1] | Ever | Since 2020 |
|---|---|---|---|---|---|---|---|---|
| International | PISA/OECD | 2000 | 2018 | 8 | 86 | Yes | 208,000 | 17,300 |
| | TIMSS/IEA | 1995 | 2019 | 7 | 64 | Yes | 24,600 | 2,650 |
| | PIRLS/IEA | 2001 | 2016 | 5 | 61 | Yes | 14,900 | 1,920 |
| | EGRA[2] | 2013 | 2019 | country level | country level | No | 7,700 | 904 |
| Regional | PASEC[3] | 1994 | 2019 | 6 | 15 | No | 1,810 | 207 |
| | SEACMEQ | 1995 | 2014 | 4 | 15 | No | 40 | 15 |
| | ERCE[3] | 1992 | 2019 | 4 | 18 | Yes | 6,060 | 833 |
| | SEA-PLM | 2019 | 2019 | 1 | 5 | Yes | 53 | 20 |
| National | NAEP / USA | 1992 | 2019 | 5 | 1 | Yes | 47,700 | 4,960 |
| | Prova Brasil / Brazil[4] | 1995 | 2019 | 13 | 1 | Yes | 135,000 | 16,300 |
| | Aristas / Uruguay[5] | 1996 | 2018 | 9 | 1 | No | 11,700 | 2,010 |
| | SIMCE/Chile[6] | 1988 | 2019 | 32 | 1 | No | 7,510 | 1,150 |

*Notes: (1)* Microdata Available for Download (no permission required); (2) the EGRA assessment framework does not necessarily ensure temporal comparability within country; and was not designed for cross country comparability; (3) only the last two rounds are temporally comparable; (4) Sample based between 1996 and 2005, and Census based from 2007 to 2019; (5) This assessment changed name, we only searched; and , (6) Census based.  Results extract from Google Scholar as of 10/21/2021.

Advocacy from key educational stakeholders, especially parents and communities, can empower communities with better data to demand more from their education systems, and encourage schools, managers, and school governing bodies to use tools to report learning at the community level. The way data are disseminated needs to be rethought. Instead of just releasing the results, ready-to-go analysis should be tailored to different types of users—from policy makers to teachers looking to reinforce their pedagogical approaches and to parents who want to understand how their child is progressing in relation to curricular expectations, and to work with their school to help reinforce their child's learning. In short, the potential of learning assessment data can be tapped by making it more accessible and usable for different actors, informing and empowering them to improve learning outcomes.

**Principle 4. Ensure data is relevant for decision-making**

Assessments must be **relevant for monitoring progress in order to inform decision-making**. The assessment results must be comparable, which means that questions have to be designed at the same level of difficulty across time, and administered to students at similar grades or ages. Moreover, it is critical that assessment must be repeated with sufficient frequency or a pre-determined regularity. Otherwise, it might not be possible to track progress, generate sufficient information for decision-making, and create accountability at the system level. Assessment timelines should also align with a country's electoral or fiscal cycle. Ideally, countries should have at least two temporally comparable measures of learning within either cycle, which can enable stakeholders to regularly account for effectiveness of current educational systems.

To ensure that assessments can accurately monitor progress for decision making, data also must be **internationally comparable** for benchmarking. Every country ought to have an assessment that in one way or another was designed for, or can be used for, international comparability—a commitment in the SDG process (not just in education). In some cases where international comparability was not built into the data collection's design, they can become internationally comparable using some of the tools and processes made available through the LDC. Even if the results may become lower compared to an original national assessment, the courage to measure against common standards can help win respect for the ministry of education with the ministry of finance, cabinet, or legislature, among other stakeholders, and legitimize reform efforts. For example, in Peru, the country's last-place ranking on PISA's global results was used to justify existing reform efforts on teacher policy, national assessments, infrastructure, and education funding.[9]

**Box 2. Data for better lives**

**Forge a new social contract for data.** For data to realize its potential to transform lives, a new social contract for data is needed. Such a contract would enable the use and reuse of data to create economic and social value, while ensuring equitable access to the value realized, as well as fostering participants' trust that they will not be harmed by data misuse. Renewed efforts are required to improve data governance domestically, as well as through closer international cooperation.

**Increase data use and reuse to realize greater value.** Increasing data access to more users through open data, interoperability standards and data sharing initiatives, for example, increases the potential of data for positive development impacts. Combining these data with traditional sources such as censuses, national surveys, government administrative data, and data produced by civil society organizations could help fill data gaps, provide timelier and finer-scale assessments of programs and policies, and serve public policy needs. Realizing this increased value calls for changing both mindsets and frameworks guiding data use.

**Create more equitable access to the benefits of data.** Major inequities in the ability to produce, utilize, and profit from data can be found across both rich and poor countries and among the rich and poor people within them. Data systems for public and private intent tend to exclude poor people, and statistical capacity and data literacy remain limited in poor countries. Many lower-income countries lack the data infrastructure and capacity needed to both produce and analyze their data in a timely and secure way.

**Foster trust through safeguards that protect people from the harm of data misuse.** The more that data are reused, the greater is the risk of data misuse, evident in growing concerns. Addressing these concerns calls for greater regulation of personal data grounded in a human rights framework, supported by policies that secure both people and the data systems they depend on.

**Work toward an integrated national data system (INDS).** Although a new social contract can rebalance and reset the rules of the game for data governance, implementing this vision further calls for an INDS that allows the flow of data among a wide array of users in a way that facilitates safe use and reuse of data. A well-functioning INDS explicitly builds data production, protection, exchange, and use into planning and decision-making and actively integrates the various stakeholders—individuals, civil society, academia, and the public and private sectors— into the data life cycle and into the governance structures of the system. Achieving a well-functioning INDS requires proper financing and incentives to produce, protect, and share data. Greater investment in physical and human capital is needed to improve data governance, specialized analytical and data security skills, as well as data literacy of the general public. Dependent on starting points, countries will need to work gradually toward this goal.

Source: World Development Report 2021.

## 4. Shared Processes

This section discusses processes countries and development partners should use to collect data under the three pillars. The LDC's approach is a comprehensive package to strengthen a country's measurement ecosystem. The package includes measuring learning across the three pillars by supporting a capacity building process for existing data collection strategies to modify their periodicity over the next decade, while countries implementing new measurement efforts are supported to develop and evolve toward that periodicity. Building capacity at the national level will support countries to adopt flexible strategies to measure learning with country constraints and contexts in mind. LDC partners collaborated on identifying three main groups of processes, namely how to measure, build national capacity, and what global and regional public goods need to be used and further developed.

The three processes are relevant for all three main pillars of the LDC: (1) measure learning progress (using large scale assessments); (2) measure the drivers of learning (using integration of different data sources including household, school, and administrative surveys); and (3) understand classroom practice (using classroom assessments). The processes in each one of these pillars are currently at different stages of maturity and development. While in the case of Pillar 1, we are on "scale-up" phase with some adjustments to the tools and methodologies, in the case of Pillars 2 and 3, we are at an earlier maturity phase. Such differences will reflect the balance of resources need for in-country implementation and technical assistance versus the development of public goods. The following describes the measurement tools and approaches for each pillar.

**Table 2. Level of Maturity of the Three LDC pillars**

| Pillars | How to Measure: Tools and Approaches | How to Build National Capacity | What Global and Regional Public Goods Exist |
|---|---|---|---|
| 1. Measure learning progress (using large-scale assessments) | High | Low-Average | Average-High |
| 2. Measure the drivers of learning (using integration of different data sources including household surveys) | High | Low-Average | Low-Average |
| 3. Understand classroom practices (using classroom assessments) | Average | Average-High | Low |

***How to Measure: Tools and Approaches***

***Pillar 1. Measuring Learning Progress*** *Using* ***Large Scale Assessments***

The LDC aims to support countries to measure learning progress using large scale assessments in at least two domains at least twice (over five years) and offers a flexible approach. Based on existing country capacities, countries have two options for implementing a large-scale assessment. One is to obtain learning data through national assessments or population-based surveys (such as the MICS Foundational Learning Module or citizen-led assessments). A second is to participate in cross-national assessments, either global or regional, though regional assessments are more prevalent at the primary level, and mostly at the end of primary.

National Learning Assessments

Implementing national learning assessments has the advantage of helping to strengthen national assessment capacity and providing a more direct link to the classroom. However, national assessments need to align with global benchmarks and be comparable over time in order to monitor progress. To develop processes that align reporting against SDG benchmarks, the Global Alliance to Monitor Learning (GAML) and the Technical Cooperation Group (TCG) for Education 2030 have facilitated the development of global public goods, the above-mentioned GPF and MPL—and created methodologies to link different national and regional assessments to a global standard specified under the GPF. These methodologies have allowed countries to use their national assessments to compare progress across global benchmarks and report results against SDG 4.1.1. Initiatives such as the Rosetta Stone and the MILO project statistically link global and regional assessments (such as IEA's PIRLS and TIMSS with ERCE and PASEC) and regional

and national assessments (such as PASEC 2019 and Kenya's National Assessment System for Monitoring Learner Achievement and Zambia's National Assessment Programme). To date, only a few national assessments have been officially linked to a global scale. More can be done to improve global comparability across existing national assessments. For example, Policy Linking Toolkit allows a country to specify the national assessment levels aligned to international benchmarks, while encouraging reflection on existing assessments and tools, and building system capacity. Bangladesh, Ghana, India, Lesotho, and Nigeria (and soon Cambodia, Nepal, and Zambia) are among the countries that have engaged in policy linking exercises.

## Cross-national learning assessment

Cross-national assessments have the advantage of generating comparability across countries and often over time. But the cost of participation varies widely, subject to the different institutional arrangements, the restricted geographical coverage in some cases, and the different approaches to capacity development of the participants' countries.

The main features of cross-national initiatives are as follows (table 1):

- The **grade and age** differ by assessment. Some assessments test at grade 4, others at grade 5, and others at grade 6. Even among similar grades, there is variation in age and years of education that each assessment has depending on the structure of the educational system (in some countries primary education comprises four grades, in others five grades, with the modal duration being six grades).
- **Subjects and numbers of grades tested** varies by test. Some regional assessments test two grades, while others assess one. Most regional assessments test two subjects.
- **Sampled populations** vary. For instance, primary completion is around 80 percent globally, but below 50 percent in many countries where children encounter the most challenges to learning. Private school enrollment, often in unregistered schools, makes up a substantial portion of enrollment in many countries. The differences affect the comparability of different tests when results aim to represent the whole group (or cohort) of a certain age.

In the last five years, there have been many developments in the tools to measure learning and align methods to global scales, such as the GPF, expanding the options for reporting on SDG indicator 4.1.1. New assessment initiatives such as SEA-PLM in South East Asia, ICAN, the MICS FLM, and the SDG4.1.1.b test associated with the UIS/MILO project have created options among the tools available to measure learning. Countries have also strengthened national assessments as participation in global and regional assessments improved methods and capacity building through learning-by-doing. Indeed, regional assessments like ERCE or PASEC provide unique opportunities for countries to build national capacity.

## Frequency and design: What is good enough?

The frequency of data collection would depend on the starting point of each country, because the best strategy is fit for purpose and depends on the context. For example, a country that has never participated in a large-scale assessment and does not have a national assessment, the time between rounds may be longer while the system builds capacity and sets an enabling context for learning assessment activities. But setting a schedule is a good habit to create a culture of regular measurement. In those cases, the preferred course of action might be to join an existing international or regional assessment program to measure learning in at least two domains, at least twice, and at least over two grades (2x2x2) to enable the country to build capacity for field work, assessment design, and more importantly data use and re-use.

Overtime as the system matures, the ideal scenario would entail measuring at least two domains, every two to three years at three education levels (2x3x3). That would allow monitoring the progress of a cohort over time, by being able to compare past cohorts with current cohorts. This ideal scenario is linked to the SDG Indicator Framework, Target 4.1. Three points of measurement on average around three years in the school cycle allow different snapshots and the longitudinal follow-up of students as they progress in the educational cycle (Figure 5).

**Figure 5. Learning as cumulative and progressive process**



Source: Azevedo, J.P., Michael F. Crawford, Marta Carnelli, and Diana Goldemberg. *Forthcoming.* Setting Targets for Progress in Reducing Learning Poverty and Improving Developmental Reading Subskills: Towards an Evidence-Based Strategy to Accelerate Learning. World Bank: Washington, DC.
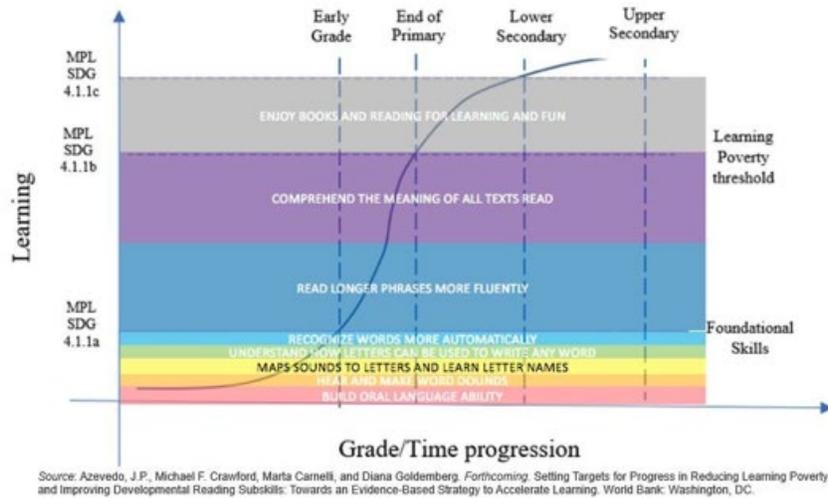
In this scenario, the 2x3x3 design would allow monitoring the progress of a cohort over time. For instance, the cohort starting in 2021/22 would complete the assessment cycle by 2031, with learning assessments in 2024, 2027, and 2030 to monitor progress though the school cycle. The design would also allow comparing 6th graders in 2024 with 6th graders in 2021 and in 2030, generating insights about education system progress. Figure 6 uses a cycle of three years as an example and illustrates alternative designs for assessment cycles. Any cycle needs to allow enough time for dissemination and uptake in the system. Furthermore, data collected under Pillar 1 needs to be aligned with data collected under Pillar 2 (on drivers of learning) and Pillar 3 (on classroom practices).

**Figure 6. Alternative designs for 2 domains, 3 levels, every 3 years**
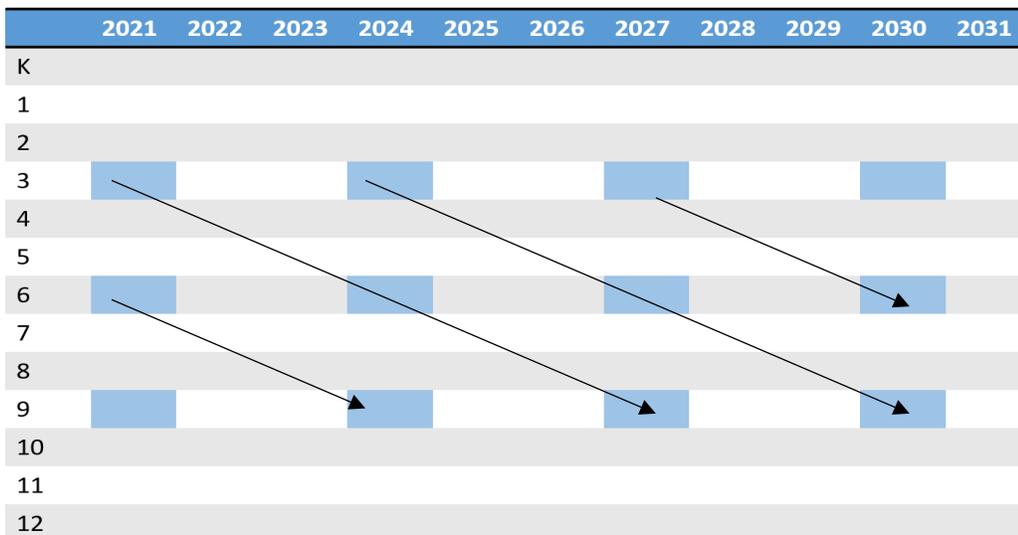
**Box 3. Integrating Transferable Skills in the Learning Data Compact**

Transferable skills[1] allow children and young people to become agile, adaptive learners and citizens equipped to navigate personal, academic, social and economic challenges. Transferable skills also support crisis-affected children to cope with trauma and build resilience in the face of adversity. Transferable skills work alongside knowledge and values to connect, reinforce, and develop other skills and build further knowledge. As such, they act as the central 'magic glue,' connecting, reinforcing, and developing other skills (foundational skills of literacy and numeracy, digital skills and job-specific skills).

Very few countries have successfully integrated transferable skills into their education systems and a major reason for this is a lack of robust approaches on how transferable skills can be measured. Measuring transferable skills will allow (i) understanding levels and distribution of transferable skills across school-age population, (ii) informing the nature and scope of required interventions to foster transferable skills teaching and learning, (iii) tracking the progress of policies and interventions designed to enhance transferable skills.

Several regional and cross-regional assessments have been or are currently developing components that go beyond SDG Indicator 4.1.1 and come closer to the orbit of 4.7 (Global Citizenship Education and Education for Sustainable Development). Regional assessments include UNICEF EAPRO and SEAMEO's SEA-PLM which feature a domain on global citizenship; UNESCO's ERCE, which in 2019 featured a module that covered socio-emotional skills, including empathy, acceptance of diversity, self-regulation and self-management; and UNICEF MENA and the World Bank's Life Skills and Citizenship Education (LSCE) initiative, including empathy, respect for diversity and self-management, among other transferable skills. The cross-regional initiatives include IEA's International Civic and Citizenship Education Study and OECD's Survey on Social and Emotional Skills as well as the PISA 2018 Innovative Domain on Global Competence.

Although the global indicator on Target 4.7 focuses on breadth of knowledge from the point of the view of the curriculum, two thematic indicators focus on global citizenship and sustainability (4.7.4) and socio-emotional learning (4.7.5). A strategy to measure indicators 4.7.4 and 4.7.5 was developed but needs updating.

The LDC will support increased collection and use of data on transferable skills by supporting a cohort of countries to:

>   1. Integrate modules on transferable skills into national assessments, particularly in countries where no information in available on transferable skills, and

>   2. Introduce or strengthen modules on transferable skills in diagnostic assessments (Pillar 2) and school and classroom-based assessments (Pillar 3).

As an overarching goal, the development and use of these modules to measure transferable skills integrated into other learning measurement efforts will allow policymakers and other stakeholders to understand the importance of transferable skills for academic learning and skills (for instance, in reading and mathematics), and the relevance of promoting transferable skills for the development of individuals and nations.

*Pillar 2: Measuring the Drivers of Learning*

Even with quality and temporally comparable data on learning outcomes from large-scale learning assessments (Pillar 1), policymakers often have insufficient information on many of the crucial factors that explains how the budget invested in education leads to the learning (or lack thereof) of students. The World Development Report 2018[10] argues that struggling education systems lack one or more of four key school-level ingredients for learning: prepared learners, quality teaching, learning-focused inputs, and the skilled management that pulls them together.

This pillar aims to strengthen countries' ability to measure drivers of (and barriers to) learning such as inputs and intermediate outcomes within the education system and household's environment following the same principles described in the earlier sections. To do so, a few options are available to countries: (i) administrative records including Education Management Information Systems (EMIS); (ii) providers surveys and classroom observation; and (iii) household surveys, with integrated measures of learning. Like Pillar 1, Pillar 2 identifies options based on current gaps in household surveys and learning modules.

**Table 3. Tools for Drivers of Learning**

| Tool | Purpose | Periodicity |
|---|---|---|
| Administrative Records and EMIS | Enable an efficient management | Annual |
| School Surveys including classroom observation | Correlate learning with inputs and outcomes within the education system | ~every 3 years |
| Household Surveys | Correlate learning with household characteristics | ~every 5 years |

**Build on what exists.** The relevant investment in measuring the drivers of learning varies from country to country depending on the maturity of existing country data systems. Here are three cases that outline what the benefits could look like for different countries.

Table 4.

| Country with low or no data infrastructure | Country with some data infrastructure | Country with advanced data infrastructure |
|---|---|---|
| Strengthening of the EMIS should be a priority. | Complementing the EMIS while strengthening it can bring value. | Even in the rare cases where good data on quality of service delivery (practices) is available, the indicators at the policy and politics levels are not generally collected, so a tool such as the GEPD can shine a light on these systemic drivers of learning. |
| Given that the process of building a strong EMIS requires years, investing in a comprehensive sample-based provider survey (such as GEPD) can fill the gap in the short term by offering much-needed information on learning bottlenecks quickly and cheaply. This can ensure that current policy decisions are evidence-based and make it possible to track progress. | Validated comprehensive provider survey can shine a light on key quality aspects of service delivery, existence and implementation of policies, and the implementation capacity and orientation of the bureaucracy, which fall outside the scope of a standard EMIS. It could also be used as auditing process for existing data. | At the practice level, the information collected can serve as an audit to existing data, while also highlighting areas where there are data gaps. |
| <mark>HH survey</mark> | Implementing validated open-source instruments and protocols with quality complementary resources can help build data-collection capacity and data infrastructure. | |
| Implementing validated open-source instruments and protocols with quality complementary resources can help build data-collection capacity and data infrastructure. | | |

**Table 5: Data Gaps in Household Surveys and Learning  Modules**

| | Global | LIC | LMIC | MIC | HIC |
|---|---|---|---|---|---|
| Total number of countries | 217 | 29 | 50 | 56 | 82 |
| Countries with a multitopic household survey used to report on SDG 1 in last 5 years (including MICS) | 81 | 19 | 34 | 21 | 7 |
| Countries that have tested learning with a household survey (MICS FLM) | 61 | 12 | 24 | 21 | 4 |

\* Note: Some countries that have tested learning with a MICS FLM have also participated in recent nationally representative citizen-led assessments, such as ASER; SDG 1 No Poverty.

## Administrative Records and Education Management Information Systems

Good administrative records, especially a good Education Management Information System (EMIS), can be used to better understand some of the drivers of (and barriers to) learning. An EMIS manages and evaluates the education system, and consists of administrative data collected from schools.  It generally collects a variety of data on child characteristics and education outcomes, school facilities and materials, and human resources. The EMIS can be used to track children's outcomes (by gender, disability, region of residence, native language, etc.), enrolment, attendance, drop-out rates, repeating grades, learning outcomes (through grades and scores on school-designed tests so generally not in standardized manner as describe under Pillar 1), school inputs (facilities, materials, human resources, budget, spending, etc.).

On delivery, the EMIS can be paper-based in which case data is collected usually once a year, often at the beginning of the school year, and limited to the classroom or school-level (no tracking of individual students). It can also be electronic allowing for the collection of more granular data (including tracking individual students or geospatial coordinates), updates throughout the year, but requires schools to have access to devices (computers/tablets) and internet.

A sustainable and well-functioning data system requires acknowledgment that the system is more than just an information technology (IT) system for storing and reporting data. A good EMIS does not stop at the data collection process. It also includes well-structured data management processes, linkages to other data sources, and a team of professionals managing it. Furthermore, an EMIS must be able answer the policy-related inquiries of decision-makers and managers, leading to on-going learning, adaptation, and improvement. It should also allow stakeholders across the education system to tap information to strengthen education system processes and practices and contribute to the resilience of the system.

**Figure 7: EMIS Complexities**[11]



---
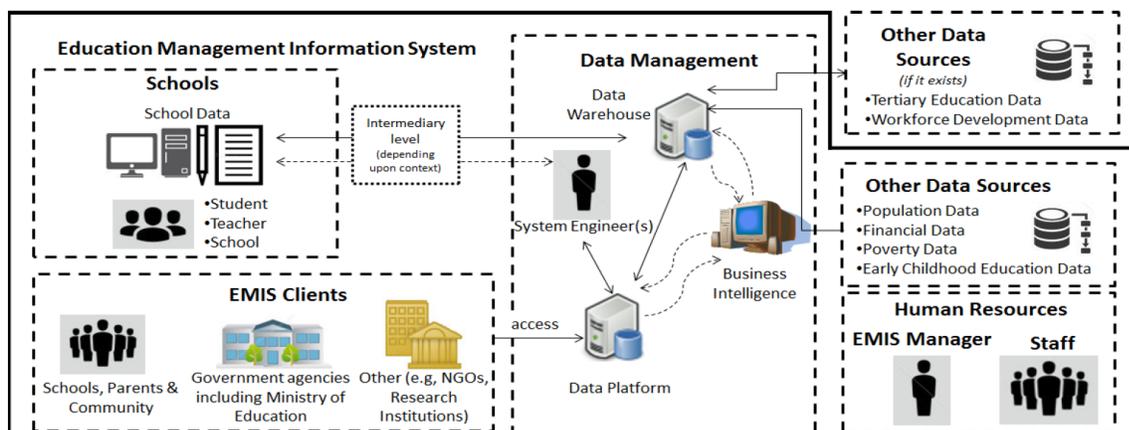
**Box 4. Open EMIS**

To put in place an electronic EMIS, countries can consider, among other options, "Open EMIS" which is an open source, electronic, and granular system created by UNESCO that can be adapted by countries. Open EMIS is being used in at least 20 countries.

---

As part of the LDC, countries are encouraged and can be supported to upgrade their EMIS so that it can be used to guide system reforms. Upgrading their information management systems can mean modifying the data collection processes to gather more relevant data more frequently. To do so, moving from a paper-based data collection method to an electronic one may be helpful but is not necessary and should be considered only if the required conditions are in place. More importantly, upgrading EMIS can also mean the use of data, by improving the relevance, quality, and timeliness of the analytics, reporting, and access to data. This likely include moving from monitoring compliance to monitoring learning which requires putting in place a team of experts in charge of carrying-out predictive analysis (of risks and opportunities), cross-data analysis (from different services and levels of education), evaluating impacts of implemented policies and programs, etc.

## School Surveys and Classroom Observations Tools

An EMIS can capture some indicators related to four key school-level ingredients for learning but not all as some may require methods not easily implemented on a routine and census basis and some may need to be collected by administrators independent of the system. And problems are not necessarily limited schools. Deficiencies in service delivery are typically signs of deeper systemic problems, driven by policies that are not well designed or implemented to promote learning, and these misalignments in turn may reflect problems caused by unhealthy politics or a lack of bureaucratic capacity. Existing instruments are designed for measuring specific pieces of the system but rare are those pulling together data on all areas of service delivery (practices), policies, and politics in a cost-effective and policy-relevant manner. This makes it difficult for a country to get a full picture of system quality without implementing many of these surveys at the same time. And because they typically tend to be quite detailed, it is expensive and time-consuming to implement all these instruments separately in a country. Furthermore, the instruction and interactions of teachers and students are strong determinants of students' learning and readiness for schooling.

School surveys, which include classroom observations that gauge how teachers interact with students, can connect administrative data and outcomes with teaching practices on the ground. School and classroom level information allows analysis to address specific and differential drivers of learning. Countries are recommended to select and administer survey tools best suited to answer relevant questions and inform decision-making and policies. There is a wealth of instruments available for collecting data on education quality at various levels. These include nationally designed school-based surveys, the World Bank Service Delivery Indicator (SDI) survey, classroom observations tools such as Stallings, CLASS, TEACH, etc., SABER policy reviews, bureaucracy lab exercises, etc. Good surveys to measure drivers of learning should capture indicators that are most likely (based on current global knowledge) to influence learning, be rolled-out in a sample designed to be representative nationally while allowing for meaningful and valid disaggregation of results, reach relevant respondents with minimum attrition, be cost-effective, be repeated every few years in temporarily comparable manner to capture change, include plans and skilled staff for analyzing and reporting on the collected data. There is value in tools that include modules measuring the quality of practices within the classroom. Classroom observations have been recognized as one of the most direct ways to measure teaching quality, since they focus on the observable behaviors exhibited by the teacher within the classroom, therefore providing a direct quality measure of the interaction between teachers and students. There is also value in using tools goes beyond measuring practices (or service delivery) by also assessing policies and politics around learning.

**Box 5. Global Education Policy Dashboard**

The World Bank, with support from the Bill and Melinda Gates Foundation, the UK's Foreign, Commonwealth, and Development Office (FCDO), and the Government of Japan, developed the *Global Education Policy Dashboard (GEPD),* which draws on existing instruments, but streamlines them by using only the pieces that are most relevant for measuring key determinants of learning. This streamlining, combined with the unified fieldwork operation, substantially cuts the cost of data collection while providing a comprehensive view of the system.

Data collection. The GEPD uses three data collection instruments:
- *School survey:* This survey is carried out in a representative random national sample of schools (200-300 schools); it takes 4 hours to complete per school. It includes 8 modules: School Information, Roster, Principal Questionnaire, Teacher Questionnaire, Classroom Observation, and Assessments for Teachers and for 1st- and 4th-Grade Students.
- *Survey of public officials:* This survey uses 45-minute interviews with education officials (around 200) at the central and subnational levels. The sample of surveyed officials is representative at the targeted levels.
- *Policy survey:* This survey involves a legislative review conducted by a senior consultant in each country, drawing on their knowledge to identify the country's choices in policy areas highlighted in the dashboard.

Furthermore, for some indicators, existing data sources (e.g. Multiple Indicator Cluster Surveys) are leveraged. Using this data, the GEPD reports on 39 indicators that, operationalizing the World Development Report 2018 framework, provide a snapshot of how the education system is working. It includes information learner readiness (grade 1), learning outcomes (grade 4), teaching practices, school participation, school management, inputs and infrastructure. In addition, it measures deeper systemic drivers: the policies and politics that determine the quality of service delivery. In addition to taking a comprehensive approach, it incorporates the latest research on education.

Open source and free resources. GEPD tools and resources are free and open source. These include survey instruments in multiple languages, programming of these instruments in Survey Solutions which allows for tablet-based data collection, user manual, resources needed for enumerator training, complementary step-by-step guides for the field staff, Terms of Reference (ToRs).

Data use. The GEPD also provides a suite of resources and the automation of important steps which enable the analysis process to require very little staff time, saving both costs and time to delivery. Aggregate-level information is reported on the GEPD website through a user-friendly interface that allows users to look at the data in a visually appealing way, track progress over time on each indicator, and do country comparisons. The platform offers features like instant regression analysis and application of statistical weights.

Example of GEPD implementation. Peru was the first country where the GEPD was implemented. Although Peru has a vast amount of data available, the GEPD adds value by providing a streamlined set of indicators on factors that have been shown to matter for learning, supported by a robust conceptual framework, literature review, and consultations with experts on the subject areas and on measurement. The dashboard also provides data analysis and reports on important indicators. Thus, the dashboard data helps focus on what matters for learning

## Household Surveys with Integrated Measurement of Learning

While reporting on SDG 4.1.1, which measures minimum proficiency in reading and mathematics at different grade levels, can be achieved using large-scale assessments (either international or national) for data collection (Pillar 1), other SDG indicators, such as 4.2.1, which measures early childhood development outcomes, 4.5.1, which measures gender, rural/urban, wealth and other parity in education outcomes, and 4.6.1, which focuses on the literacy and numeracy skills of youth and adults, would require

data collection to happen outside of the classroom. Multi-topic household surveys that are regularly carried out in most developing countries around the world offer a possible vehicle to collect such data, and other relevant learning data, at scale.

The approach requires identifying a main survey instrument and to integrate to this instrument one or several learning data modules. Ideally, the main survey instrument would be a multi-purpose household survey such as LSMS, MICS or DHS which typically covers quite a large array of topics including demographics, dwelling conditions, full consumption/expenditures, education, health, labor, social protection, agriculture and food security, etc. These data will be key to understand the drivers and impacts of learning outcomes from the household perspective. There might also be some additional variables that countries may want to introduce to the survey instrument depending on needs.

Learning measurement in the context of household surveys should rely on established instruments that have been validated in multiple and comparable contexts and are designed to measure learning at the targeted age/grade. The specific age/grade would depend on countries' objectives and existing data gaps. Examples of such instruments include the ECDI, the MELQO and AIM-ECD, the FLM, the FLE and mini-LANA and mini-MILO, the PISA-HSM, and the mini-LAMP (see below Box 6 about the initiative of integrating several of these to cover all possible age groups).

While household surveys cannot substitute for international or national large-scale assessments of in-school students (Pillar 1) and cannot substitute for in-depth data collection on drivers of learning within the practices, policies and politics of education systems (Pillar 2), there are several reasons why such instruments could be complemented by measures of skills or student learning via household surveys. Multi-topic household surveys with learning assessment modules would allow to:
- *Capture learning of a representative sample of out-of-school population*, be it due to age (i.e. pre-school or post-school) or due to education status (i.e. school-age dropouts), including in fragile and conflict affected contexts or in refugee populations especially those living in camps.
- *Correlate measured skills or competencies with a diverse set of individual-, household-, and community-level variables* (collected as part of these surveys) which can provide significant scope for better understanding the factors that could be driving learning outcomes, beyond

factors operating within the education system. In particular, this would allow a better understanding the drivers of SDG4 outcomes by relying on the richness of the variables

---

**Box 6. Lifelong learning**

The World Bank (WB), the United Nations Educational, Scientific and Cultural Organization (UNESCO) Institute of Statistics (UIS), the UNESCO Institute for Lifelong Learning (UIL), United Nations Children's Fund (UNICEF), International Association for the Evaluation of Educational Achievement (IEA), and the Organization for Economic Co-operation and Development (OECD) are collaborating on a work program that would enable countries to administer short learning assessments as part of household surveys that could potentially cover all possible age groups.

Data collection. The approach will integrate valid and reliable established tools that measure the abilities of individuals to meet certain minimum proficiency thresholds on competences, learning and skills which have already been (or can be prepared to be) implemented as part of a household survey. Each learning module would have a maximum length of 30 minutes per respondent and be able to discriminate around relevant proficiency thresholds aligned with the SDGs. The learning measurement modules would be conducted either in the same visit or in a sequenced follow-up visit fully integrated to the main survey.

The initiative is currently at the design stage and will be piloted in select middle- and low-income countries, building opportunistically on a planned national- or subnational-representative household survey. The proposal is to collect learning data using six existing age/grade-specific instruments (see figure below).

**Figure 8. Age Ranges of Current Assessment Tools**
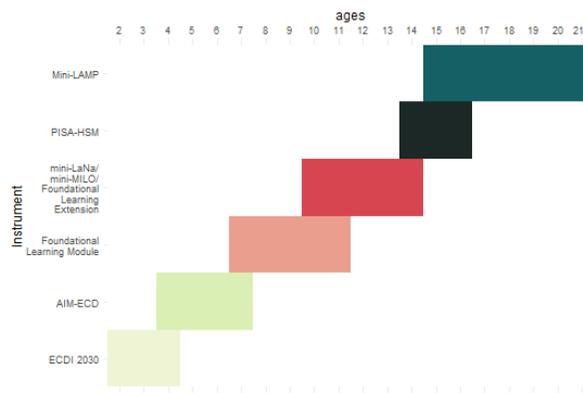


Open source and free resources. The modules will be made available, free of charge, to any organizations conducting household surveys, so long as they abide by licensing rules and guarantee that items remain confidential, and users will be encouraged to in turn make microdata available publicly. Guidelines for administered through different delivery modalities (one-on-one in person interview, self-assessment, and phone-based) and modes (paper-based, computer-based) will be developed and made available.

Data use. The data will help to identify the characteristics and risk factors of dropouts, testing the hypothesis on whether the learning levels of the child (and household members) are a significant risk factor for student dropout and enabling the policies to be more tailored and better targeted, and thus more effective and efficient. The data also serve as a source for a comprehensive situation analysis by identifying the learning poor, and children left behind from the education system.  It allows identifying at what point children dropped out from the education system. In addition, where population census data, administrative records, and geospatial information is available, we would be able to use small area estimation techniques (i.e. poverty mapping methodology) to generate more granular maps and spatially disaggregated data to further diagnose the issues related to dropout and improve targeting. Finally, the initiative would produce estimates of the magnitude of learning gaps between in-school and out-of-school students with similar characteristics, which can be used as a powerful advocacy tool. All data would be fully documented and made publicly available for anyone to use.

collected through the household survey, enabling analysis of intra-household and inter-household variation as well as intergenerational dynamics. It would also allow to link outcomes on SDG4 to other SDG such as engagement of child labour (SDG8.7), disability status (SDG4.5.1), mother's education, parental engagement in education and stimulus and positive home environment (SDG 4.2.3).

• *Apply different administration methods (e.g. caregiver reports for young children and direct assessment for older children and adults) and triangulate* direct assessment of an individual's skills or competences with other-reports by household members (e.g., direct assessment of school-age children and caregiver reports) to derive a more comprehensive measure of these skills and competences.

• *Build such capacity while temporarily filling an important SDG 4 learning assessment data gap* at relatively low cost in contexts where there is insufficient national capacity or resources to carry out large-scale assessments of in-school students

• *Link data on learning at the household level to administrative and service delivery data* (on health, education, transport, etc.).

## Citizen led learning assessment (UWEZO, ASER)

Citizen-led assessments are household-based assessments implemented by local organizations, and can be used to understand drivers of learning. Modules typically include how children learn through access both at home and school, and cover children both in-school and out-of-school, and rely on simple tools that are easy to implement. These assessments have commonly been implemented at low income and highly populated areas, and do a good job at raising public awareness about low learning levels. However, there is little evidence of citizen-led assessments leading to improvements in learning quality.[12] While citizen-led assessments are not meant to be diagnostic and do not feed into individualized feedback to children, they can be used as effective advocacy tools to raise awareness about the learning crisis.

### Pillar 3. Understand Classroom Practices: Aligning Classroom Assessment with National Learning Assessments and/or the Global Proficiency Framework

One of the central challenges affecting education systems is the inability to collect and use data at school and classroom levels to inform practices. Effective policies require contextualized approaches which are not only country, but also school and classroom specific. The assessments and data produced by activities in the first two pillars provide critical information on system performance and diagnosis of the factors contributing to or inhibiting learning. This pillar is focused on two types of data, namely:

1. **Classroom assessments for learning** focused on giving teachers the data they require to inform teaching practices including targeted instructions and grouping of children.

2. **Classroom and school practice measurements** to understand applications and processes that influence learning

The third pillar seeks to embed school and classroom assessments in education systems in order to inform how learners are supported in daily practice and align school and classroom-based assessments with national learning assessments and other system data. It will invest in supporting country systems to adopt appropriate approaches and tools, build capacity of education personnel, and support implementation of assessment and use of data. As with the rest of the LDC, this component will:

• Build on existing instruments and capacity development process;
• Integrate tools and approaches into teacher, curriculum, and school planning processes; and,

## Classroom assessments for learning

Classroom assessments should align with and fit within the broader assessment system/ecosystem (i.e., alignment with Pillar 1 and Pillar 2). Therefore, this component of Pillar 3 will focus on strengthening system capacity to embed classroom assessments in the following ways:

- Support the design and development of context-relevant classroom assessments;
- Strengthen and support regular use of classroom assessments – with strong component on teacher training and ongoing support;
- Integrate classroom assessments in curricular frameworks and teacher training and development; and,
- Improve equity in learning outcomes as classroom assessments help to ensure those who are traditionally left behind from *summative dependent assessment of learning* are provided support on an ongoing basis.

Integrating classroom assessments in education systems means that they are not only reflected in education plans and policies, curricular frameworks and teacher training and development, but also actively used in the classrooms to inform teaching. At the classroom level, formative and summative assessments help improve quality of instruction and refine pedagogical approaches by giving information on students' learning progress. Combined they allow classrooms to have the data required to support learning but need to be balanced and allow for all approaches to cyclically inform and change instructions. They also help students, teachers, supervisors and school management hold one another accountable for achieving learning goals.

- To inform instruction and pedagogy;
- To ensure accountability towards learning goals;
- Helps the teacher and school to understand how children are performing on an ongoing basis; and
- Provides opportunities for corrective action **during the learning process** not only after.

The success of classroom assessments hinges on teacher capacity and motivation. Therefore, it is important that teachers are well-equipped to assess children's learning and tailor instruction and pedagogy accordingly. Support at all stages of the teacher's profession life cycle – pre-service, in-service and continuous professional development. Effective approaches also ensure the links to other areas including classroom management.

## Instruments/tools used in assessment for learning

Despite the need for *assessment for learning* to be localized and specific to context, there are a range of tools to support school based formative assessments. The LDC will identify and support countries in adapting tools, building capacity for their use, and support for implementation. The tools identified below is not exhaustive but gives a sense of the range of options available and on which the LDC will build.

*Formative assessment tools:* UNICEF and Pratham developed formative assessment tools for reading and math which are based on the Annual Status of Education Report (ASER).  These tools are geared towards supporting teachers to use in the classroom and are designed for use with minimal additional support. The tools include a users' guide and a range of tools and videos to help teachers in applying the tools.

***Citizen-led assessments:*** This is another set of simple tools based on the ASER to assess the learning level of students from, especially those from marginalized groups. These tools are available and could be useful for country application in classrooms.

**Teacher capacity building on formative assessment**: A number of tools to support teachers in doing formative assessments are in use globally. Each of the tools listed in this section is accompanied by a teacher support guide. One of the more recent guides for teachers' professional development is the Formative Assessment for Quality, Inclusive Digital and Distance Learning during and beyond the COVID-19 Pandemic.

***Assessment informed [instruction](#)*** *(Tools for Classroom):* This is a guide to help school systems implement assessment at scale to support FLN. It is designed to clarify assessment related issues and help systems decide how to approach assessment for learning – it complements the other tools presented in this pillar.

[All Children Learning Assessment Platform (ACLAP)](#): the ACLAP is an online hub through which education primarily in South East Asia track progress on their assessments and access a range of assessment tools

### Assessment of schools and classrooms
Understanding what's happening in schools and classrooms is key to developing deep knowledge of the so-called *class effect* that large-scale assessment identify but are not able to explain. This can be done through research on practices and processes in schools and classrooms to inform effectiveness of teaching practices, teachers' presence as well as effective participation and engagement of and with pupils, including the most in need. This data can be integrated with classroom assessment and national assessment to inform decision making related to teaching practices, teachers' behavior but also necessary curricular reform, to enhance the learning process.

To strengthen understanding of school and classroom contexts and practices, the LDC will build on initiatives aiming at national capacity development effective national ownership of the methodologies implemented, and for which methodological guidance are already available in the public domain.

### Tools for assessment of schools and classrooms
The range of tools available at the global level for this purpose is also quite varied. Some are focused on specific aspects of schools such as teacher supervision, teaching practices, or school infrastructure. The LDC will focus tools that are comprehensive (whole school) or capture data across multiple domains which are known to be drivers of learning. This is therefore closely linked to the data produced under Pillar 2 both in terms of use of the data and developing additional tools for schools and education systems to use to assess performance of schools and classrooms.

***School profiles and report cards:*** *Data from* assessment of schools and classrooms as well as other sources identified in Pillars 1 and 2 can be used to develop the school profiles and school report cards. These can provide general overview of the performance of the schools and factors affecting/driving learning. These are useful to help the school and the local district and can drive the types of supported required from these levels for the schools. Add relevant country examples

***Data must speak (DMS) positive deviance research:*** This component of the UNICEF-led DMS global initiative, is a multi-year research being implemented in fifteen countries in Africa, Asia and Latin America. The DMS research aims at: (1) Identifying positive deviant schools i.e., schools that are outperforming others

even though they operate in similar context and with equivalent resources; (2) Understanding what are the positive deviant behaviors and practices implemented in those schools – "What Works" and; (3) Investigating actionable system, school, and community levers to scale those positive deviant behaviors and practices to less-performing schools in a given context – "How To".

**Figure 9.**



Stage 0
Global Methodological Review

| Stage 1 | Stage 2 | Stage 3 | Stage 4 | Stage 5 |
|---|---|---|---|---|
| Analysis of resources and context associated with school performance *[Quantitative Research]* | Identification of Positive Deviant Schools and School Typology *[Quantitative Research]* | Understanding School-level Positive Deviant Behaviors/Practices *[Behavioral Science]* | Investigating Levers for Optimum Scale *[Participatory Implementation Research & Scaling Science]* | Country-level Knowledge Use & Global Mobilization |

***Early learning assessment module:*** The early learning assessment (ELA) module, undertaken at the beginning of primary education, assesses young children skills (socio emotional, cognitive skills, psychomotricity, language, relationship) and adequate children preparedness to attend primary education. This assessment will: evaluate children's preparedness in relation to their background, therefore informing value addition of preschool attendance; influence preschool duration and curriculum reforms; Inform weakness that early primary curriculum should be able to address, by an enhanced focus (eg. language of instructions). Early Learning assessment was implemented in xx countries in West and Central Africa and used to inform critical policy decisions[13].

***Classroom observations:*** Classroom observation can be used in diverse ways including to assess 1) the effectiveness of teaching practices and their alignment with the curriculum and national guidance and 2) pupils' participation and effective engagement in class[14]. Some specific tools??

***Time to teach:*** Time to Teach[15] identifies factors affecting teacher attendance and uses this evidence to inform the design and implementation of teacher policies. Drawing from both quantitative and qualitative data on a range of topics—from motivation to retention—this research aims to identify solutions for sustainable change.

### How to Build Capacity: A Country-Centric Approach to Capacity Development

Building capacity in-country is critical for a sustainable measurement system that uses learning data to drive improvement. This component is difficult to cost because it depends specifically on country context and measurement history, such as whether the country has recently participated in a cross-national assessment, among other aspects. This component should include all the support countries need to strengthen national assessments, including support to improve the quality of assessment instruments, strengthen assessment procedures, build staff capacity, and create solid research networks. **This could**

**involve at least the same level of investment as the pure measurement component,** because without it the pure measurement component will be neither very useful nor very sustainable**.**


*What Global and Regional Public Goods to Use and What to Further Develop*

Building on the successes of the last five years, a close collaboration and partnership approach has helped to align learning assessments globally (GPF and MPL). Comparable reporting that uses all available information (policy linking, subject-based and item-based linking) and innovations (FLM, MILO) require the following steps.

**Step 1** is to take the low-hanging fruit—**create an assessment repository by compiling and cataloging global and regional resources for capacity building as well as guidelines and instruments for implementation**, including item banks, that countries can draw on and use to fill implementation gaps. Building an assessment repository would make it easier to conduct and expand assessments and reduce the cost of developing, implementing, and using assessment instruments. These guidelines and instruments would facilitate the development administration, and reporting of assessments. Such a repository could also include example items (actual items or item constructs) for linking national assessments to global standards, tips and examples for ex-post linking, including case studies, and other ideas.

**Step 2** is to **strengthen regional assessments** given their critical role in building capacity and generating a unique peer learning and community of practice. Regional assessments present a unique platform for countries that share common features. They have more of a focus on primary schooling and on the foundational learning stages compared to international assessments and are key to uplifting learning levels in low and lower middle-income countries.

**Step 3** is to **mobilize additional funding to support innovative public goods** that include new alignment and linking methodologies and studies, and the development of a global item bank that the UNESCO Institute for Statistics and World Bank Group are leading. Or the development of a socio-emotional skill module to be implemented in household surveys. These products could feed the assessment repository mentioned in the first step. This last element is particular important for pillars 2 and 3.


## 5.  Shared Accountability

*LDC Governance Arrangement*

To foster accountability in its operations, the LDC proposes a shared governance arrangement, with three main elements:

1) A **Technical Management Committee**, comprising staff from the three organizations (UNESCO, UNICEF, and the World Bank), as well as other development partners, who meet regularly to coordinate the development of public goods.
    a. The technical management committee will be responsible for the day-to-day management of the programme and implementation according to the implementation plan, budget and results framework. The committee will produce the quarterly report to be approved by the Steering Committee and report back at the regular progress meetings.
    b. As part of their responsibility the members of the Technical Management Committee will closely monitor implementation and liaise directly with the technical specialists that are

leading the specific technical activities. The members will also ensure close coordination among the three consortium members.

    c. The technical teams liaise directly with co-implementing partners, country-based grant agents, particularly UNICEF, the grant agent at country level managing the initial catalytic funding provided by GPE.

2) A **Steering Committee** convened by the three organizations. The main responsibility of this group, with key stakeholders, will be to meet with predetermined regularity (such as bi-annually) to review the progress of the LDC. It includes education leads from ADG Education (UNESCO), Education Global Practice (World Bank), and UNICEF Education.

    a. The steering committee will provide overall direction for the programme against the agreed implementation plan, budget and results framework. It will provide an accountability function for progress against the plan, budget and results and will agree updates that form basis of the regular progress meetings. Six-monthly reports will also be agreed by the steering committee.

3) A **Public Progress Dashboard**, which will be updated regularly, mapping the activities and progress of different LDC related activities nationally, regionally, and globally.

**Figure 10: Members of Steering Committee and Technical Management Committee**
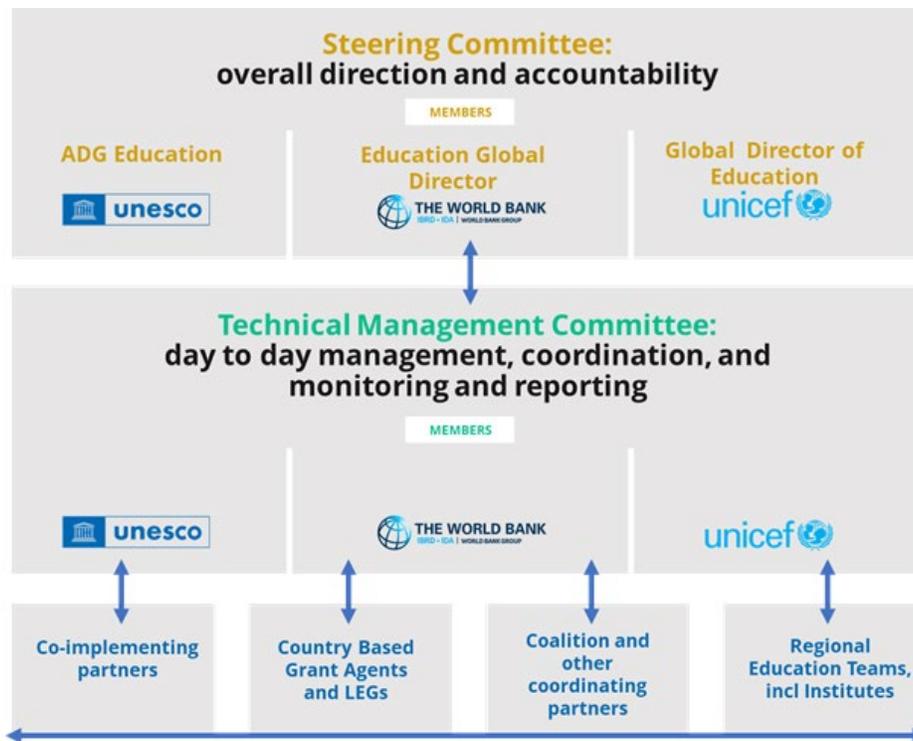


Daily management and governance, to be discussed, would focus on prioritizing the following areas:

- Identifying and addressing gaps in existing support at the country level.
- Effectively brokering support between development partners and countries.
- Conducting strategic dialogues among development partners, countries, and other relevant stakeholders.
- Tracking actions against principles and commitments.

- Monitoring activities, finding synergies, and avoiding duplication of efforts

***Partners and Country Commitments***

All development partners participating in the LDC will commit to agree on priority countries and the global and regional public goods that need to be supported, to disburse funds accordingly, and to submit minimal tracking data once a year to be shared publicly. Those commitments will follow the shared principles, processes and accountability set by the LDC. The contributions of each development partner will be registered as part of the common accountability dashboard mentioned earlier and will be documented in the bi-annual progress report. The tracking is meant to detect whether funding is funneled according to agreed priorities, whether duplications of effort and initiatives are avoided, and whether there is close collaboration among development partners around long term goals to facilitate sustainability of the assessment system.[16]

External funding through existing lending mechanisms and/or grants are envisioned as a complement to domestic resources, and it is expected that development partners would continue to provide support, and recipients continue to receive support, but in a more efficient, intentional, and coordinated manner.

**Figure 11. Learning Data Compact: Operational Model**
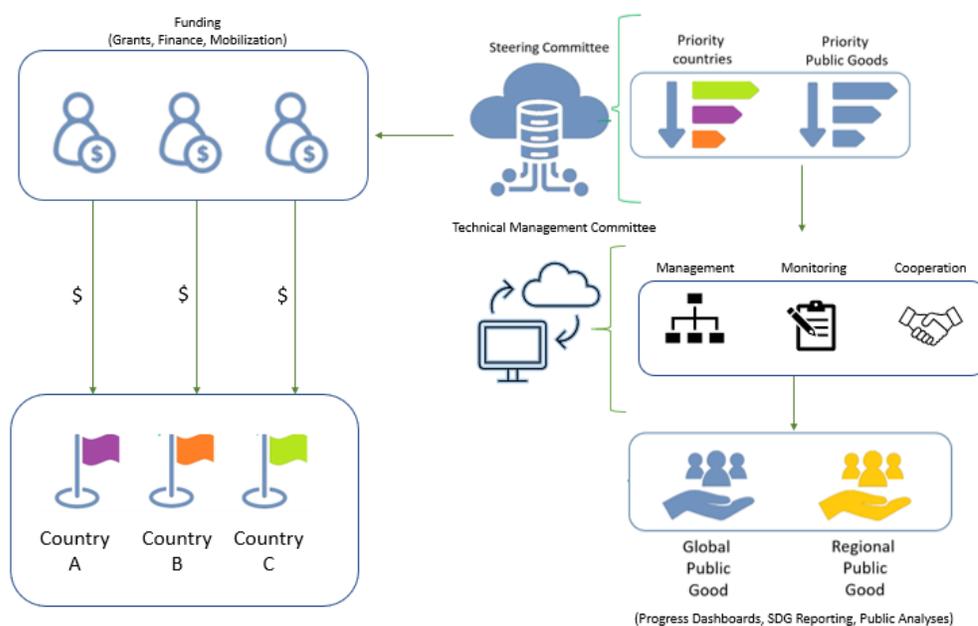


## 6. Costing the LDC

*Pillar 1: Measuring Learning Progress: Estimated Cost*

This section describes the costing model for the LDC for Pillar 1.[17] We identify total costs for countries to attain the goal of assessments for at least 2 grades, 2 subjects, and 1 round (2 x 2 x 1) under different scenarios based on assessment type.[18] We focus on the 132 countries that do not have a recent measure at 2 or more SDG 4.1.1 measurement points,[19] with a particular emphasis on the 58 IDA/Blend and 37 IBRD countries in this category.[20]

Currently, countries that are on track to attain the 2 x 2 x 1 goal have spent a total of 113 million, or around 1 million per country in the last 3 years. While other countries may follow a similar costing model, there are important differences in data gaps and needs that will vary costs and options.

Under an "intermediate" or moderate cost scenario for countries with less than two recent SDG 4.1.1 measurement points, the total cost for implementing assessments across 2 grades, 2 subjects, and 1 round is estimated at $227 million, or roughly $1.7 million per country. This scenario implies that the four assessment types (policy or student linking, national learning assessment with policy or student linking, regional assessment, and international assessment) are evenly distributed among countries that do not have any SDG measurement points over the last 7 or more years, or have SDG measurement points older than 7 years. For IDA/Blend and IBRD countries with less than two recent SDG measurement points, the total is $111 million and $94 million respectively under the scenario.

It is important to note that the "intermediate" scenario is just one way to reach the goal of 2 x 2 x 1. There are various other combinations of assessments that can be used to reach this goal, which imply different costs. For example, under a "low" cost scenario the total cost for implementing assessments across 2 grades, 2 subjects, and 1 round for countries that do not have a recent learning measure at 2 or more SDG 4.1.1 measurement points is estimated to be around $209 million, or roughly $1.6 million per country. The corresponding cost under a "high" scenario is $272 million, which is around $2.1 million per country. These alternative scenarios (low, intermediate, high, custom) are discussed in greater detail below.

This section describes the costing model for the LDC for Pillar 1.[2] We identify total costs for countries to attain the goal of assessments for at least 2 grades, 2 subjects, and 1 round (2 x 2 x 1) under different scenarios based on assessment type.[3] We focus on the 132 countries that do not have a recent measure at 2 or more SDG 4.1.1 measurement points,[4] with a particular emphasis on the 58 IDA/Blend and 37 IBRD countries in this category.[5]

Currently, countries that are on track to attain the 2 x 2 x 1 goal have spent a total of 113 million, or around 1 million per country in the last 3 years. While other countries may follow a similar costing model, there are important differences in data gaps and needs that will vary costs and options.

Under an "intermediate" or moderate cost scenario for countries with less than two recent SDG 4.1.1 measurement points, the total cost for implementing assessments across 2 grades, 2 subjects, and 1 round is estimated at $227 million, or roughly $1.7 million per country. This scenario implies that the four assessment types (policy or student linking, national learning assessment with policy or student linking, regional assessment, and international assessment) are evenly distributed among countries that do not have any SDG measurement points over the last 7 or more years, or have SDG measurement points older than 7 years. For IDA/Blend and IBRD countries with less than two recent SDG measurement points, the total is $111 million and $94 million respectively under the scenario.

It is important to note that the "intermediate" scenario is just one way to reach the goal of 2 x 2 x 1. There are various other combinations of assessments that can be used to reach this goal, which imply different costs. For example, under a "low" cost scenario the total cost for implementing assessments across 2

---

[2] This note accompanies the Excel tool containing the data and costing simulations. The Excel tool can be accessed ==here==.

[3] The cost simulations below are for one round of assessment only.

[4] SDG Indicator 4.1.1: Proportion of children and young people (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex.

[5] See the "Data and Methodology" section for details on methodology used to categorize countries into different groups, and the number of countries in each group.

grades, 2 subjects, and 1 round for countries that do not have a recent learning measure at 2 or more SDG 4.1.1 measurement points is estimated to be around $209 million, or roughly $1.6 million per country. The corresponding cost under a "high" scenario is $272 million, which is around $2.1 million per country. These alternative scenarios (low, intermediate, high, custom) are discussed in greater detail below.

Currently, countries that are on track to attain the 2 x 2 x 1 goal have spent a total of 113 million, or around 1 million per country in the last 3 years. While other countries may follow a similar costing model, there are important differences in data gaps and needs that will vary costs and options.

Under an "intermediate" or moderate cost scenario for countries with less than two recent SDG 4.1.1 measurement points, the total cost for implementing assessments across 2 grades, 2 subjects, and 1 round is estimated at $227 million, or roughly $1.7 million per country. This scenario implies that the four assessment types (policy or student linking, national learning assessment with policy or student linking, regional assessment, and international assessment) are evenly distributed among countries that do not have any SDG measurement points over the last 7 or more years, or have SDG measurement points older than 7 years. For IDA/Blend and IBRD countries with less than two recent SDG measurement points, the total is $111 million and $94 million respectively under the scenario.

It is important to note that the "intermediate" scenario is just one way to reach the goal of 2 x 2 x 1. There are various other combinations of assessments that can be used to reach this goal, which imply different costs. For example, under a "low" cost scenario the total cost for implementing assessments across 2 grades, 2 subjects, and 1 round for countries that do not have a recent learning measure at 2 or more SDG 4.1.1 measurement points is estimated to be around $209 million, or roughly $1.6 million per country. The corresponding cost under a "high" scenario is $272 million, which is around $2.1 million per country. These alternative scenarios (low, intermediate, high, custom) are discussed in greater detail below.

The four key principles of the LDC underlie the cost simulations:

**Build on what exists**

The costing exercise starts with the most recent assessments (over the past 7 years) that countries use to report on SDG 4.1.1a, b, and c. Our costing simulations allow reporting countries to build on existing assessments to achieve the 2 x 2 x 1 goal rather than having countries start from scratch. The most recent assessments that countries have used to report on SDG 4.1.1a, b, c are also used to calculate current baseline costs. The current total global baseline cost of the assessments market based on countries existing participation in assessments is roughly $213 million. For IDA/Blend and IBRD countries, the estimate is $55 million and $85 million respectively.[21] We build on these baseline costs to simulate achievement of the 2 x 2 x 1 goal.

**Allow flexibility to reach the 2 x 2 x 1 goal (and ensure alignment with country needs – not one-size-fits-all)**

The different simulations scenarios (low, intermediate, high, custom) show that countries have different options to reach the goal of 2 x 2 x 1, depending on the type of assessment chosen. Countries may choose assessments most suitable to their context and constraints, choosing different assessments to determine costs. We build four scenarios – low, intermediate, high, and custom – each implying a different proportion of countries choosing different types of assessments, with the "custom" scenario allowing the user to select two assessments of their own choice for each country. This flexibility is important to ensure that a country's chosen assessments are fit-for-purpose, while offering flexible options to attain the 2 x 2 x 1 goal. For example, a country that has a national learning assessment of sufficient quality could participate in policy linking or student linking, instead of developing a national learning assessment from scratch. Similarly, a country that is already taking part in a regional learning assessment may continue to

build on that in the future, instead of participating in other new assessments. Continued participation in the same assessment has important implications for temporal comparability.

Making careful decisions about the type of assessment is also important because the time to develop and roll out assessments may differ by assessment type. For example, policy linking on an existing national assessment or MILO may take 1 year from field work to results. For regional assessments, the time between field work and results might be 2 years. For national learning assessments, the time frame might be longer, and could involve 1 year for preparation, 1 year for field work, and 1 year to get results.

**Foster country ownership through a demand-driven approach**

The assessment options allow transparent, reliable, and quality data to ensure that they contribute to effective data use at the country-level, which in turn can spur demand. The assessments are benchmarked against GPF to support flexible capacities to obtain quality data, with collaborated efforts to allow every round of assessment data to be reused across sectors for analysis.

**Ensure data is relevant for decision-making**

The 2 x 2 x 1 goal is designed to create capacity for countries to implement the same assessment to report results over time, and hence monitor progress and inform decision-making. While countries have flexible approaches to attain the goal, efforts must also be sustained for regular measurement. Thus, they should continue to assess student learning using the same assessments implemented in the first round, so that they can strengthen existing processes to collect data and track improvements in learning over time.

## Data and Methodology

*Assessment Data*

We use the latest SDG data on assessments countries have used to report on SDG indicator 4.1.1a, b, and c over the last 7 years (2014-2020). We use this data to identify the most recent assessments that countries used to report on SDG 4.1.1a, b, and c respectively. Using this data, we divide the 217 countries in our sample into 4 main groups:

- *Group 1.* These countries have reported on 2 or more SDG measurement points through assessments that took place in the last 3 years, and are considered "on-track." Group 1 contains 85 countries (shown by the green box in Figure A3), out of which 16 are IDA/Blend and 33 are IBRD.
- *Group 2a.* These countries have reported on 2 or more SDG measurement points through assessments that took place more than 3 years ago but less than 7 years ago. Since these assessments are dated, such countries are not considered "on-track." Group 2a contains 20 countries (shown by the orange boxes in Figure A3), out of which 8 are IDA/Blend and 9 are IBRD.
- *Group 2b.* These countries have only reported on 1 SDG measurement point through assessments that took place over the last 7 years, and are not considered "on track." Group 2b contains 34 countries (shown by the yellow boxes in Figure A3), out of which 14 are IDA/Blend and 13 are IBRD.
- *Group 3.* These countries have reported on either no SDG measurement points or have reported on one or more SDG measurement point through an assessment that is older than 7 years. Essentially, all countries that are not in Group 1, 2a, or 2b fall under Group 3. Such countries are also not considered "on-track." Group 3 contains 78 countries (shown by the red boxes in Figure A3), out of which 36 are IDA/Blend and 15 are IBRD.

*Cost Data*

We use the cost data for the various assessments based on information from OECD, PISA, IEA, UIS, and Wagner (2011). This cost data was compiled by UIS.

**Table 6. Assessment Dimensions**

| Responsible Organization | Assessment | Cost for Grades | Cost for Subjects | Grade 1 | Grade 2 | Subject 1 | Subject 2 |
|---|---|---|---|---|---|---|---|
| **Global** | | | | | | | |
| International Association for the Evaluation of Educational Achievement (IEA) | PIRLS | 1 | 1 | 4 | - | Read | - |
| International Association for the Evaluation of Educational Achievement (IEA) | TIMSS | 2 | 2 | 4 | 8 | Science | Math |
| Organisation for Economic Co-operation and Development (OECD) | PISA | 1 | 2 | 7 | - | Read | Math |
| Organisation for Economic Co-operation and Development (OECD) | PISA4D | 1 | 2 | - | - | - | - |
| **Regional** | | | | | | | |
| UNESCO | ERCE/TERCE | 2 | 2 | 3 | 6 | Read | Math |
| SEACMEQ | SACMEQ | 1 | 2 | 6 | - | Read | Math |
| CONFEMEN | PASEC | 2 | 2 | 2 | 6 | Read | Math |
| SEAMEO | SEA-PLM | 1 | 2 | 5 | - | Read | Math |
| SPC | PILNA | 2 | 2 | 4 | 6 | Read | Math |
| **Other** | | | | | | | |
| UNICEF (Option 1) | MICS (No prior MICS) | 1 | 2 | any | - | Read | Math |
| UNICEF (Option 2) | MICS (Existing household survey) | 1 | 2 | any | - | Read | Math |
| UNICEF (Option 3) | MICS (Prior MICS Survey) | 1 | 2 | any | - | Read | Math |
| PAL NETWORK | PAL Network | 1 | 2 | any | - | - | - |
| National assessments and policy or student linking | Countries | 1 | 2 | any | - | Read | Math |
| National assessments without policy or student linking | Countries | 1 | 2 | any | - | Read | Math |
| Policy or student linking only | Countries | 1 | 2 | any | - | Read | Math |
| MILO | MILO | 1 | 2 | 6 | - | Read | Math |
| USAID | EGRA | 1 | 2 | 2 | - | Read | Read Second Language |

**Table 7. Data production: Cost per assessment per cycle (in US dollars, hundred thousands)**

| Assessment | National government staff | Institutional Costs and Test Fees | Fees | Test Admin | Processing and Analysis | Outreach and Dissemination | Other Costs | Total Cost |
|---|---|---|---|---|---|---|---|---|
| **Global** | | | | | | | | |
| PIRLS | 1.0 | | 2.5 | 4.2 | 1.0 | 1.0 | | 9.8 |
| TIMSS | 1.0 | | 4.5 | 4.2 | 1.0 | 1.0 | | 11.7 |
| PISA | 1.0 | | 2.5 | 4.2 | 1.0 | 1.0 | | 9.7 |
| PISA 4D | 1.0 | | 2.5 | 4.2 | 1.0 | 1.0 | | 9.7 |
| **Regional** | | | | | | | | |
| ERCE/TERCE | 1.0 | 0.3 | 2.0 | 2.9 | 2.0 | 1.5 | | 5.8 |
| SACMEQ | 1.0 | 0.7 | 1.5 | 6.7 | 1.5 | 1.5 | 1.0 | 6.7 |
| PASEC | 1.0 | 0.4 | 5.0 | 3.6 | 5.0 | 1.5 | | 8.9 |
| SEAPLM | 1.0 | 0.4 | | 4.2 | | 1.2 | 1.0 | 4.6 |
| PILNA | 1.0 | 0.4 | 0.7 | 3.5 | 0.7 | 0.8 | | 3.8 |
| **Other** | | | | | | | | |
| MICS (No prior MICS) | | 1.3 | | 1.3 | $ - | 3.5 | 2.5 | 20.1 |
| MICS (Existing household survey) | | 1.5 | | 1.5 | $ - | 3.5 | 5.0 | 6.0 |
| MICS (Prior MICS Survey) | | 1.1 | | 1.5 | $ - | 3.5 | 2.5 | 5.3 |
| PAL NETWORK | | | | | $ - | 2.5 | | 3.5 |
| Countries | 10.0 | | | | | 2.5 | | 15.3 |
| Countries | 10.0 | | | | | 2.5 | | 1.4 |
| Countries | | | | | | | | 1.8 |
| MILO | 1.0 | | 1.8 | | 1.8 | 2.5 | | 6.3 |
| EGRA | 5.0 | | | 1.3 | | 0.9 | | 7.7 |

*\*MICS is expanding its FLS module to include 2 grades (4.1.1a and b), which is expected to be implemented in 2 years*

We standardize the costs for each assessment so that they represent the cost of doing an assessment for: (i) 1 grade and 1 subject, (ii) 2 grades and 2 subjects, and (iii) only 1 grade or 1 subject. This standardization ensures comparability of costs across assessments and helps facilitate calculations underlying the simulations. For assessments that do not cover 2 grades and 2 subjects, we use an 80% discount factor (on the original total cost) to adjust for expanding the scope of the assessment to an additional grade or an additional subject. This calculation assumes that there are economies of scale that bring down the additional cost of expanding to an additional grade or additional subject. For example, since PIRLS covers one grade and one subject, we first use an 80% discount factor on the original cost to add an additional grade, and subsequently apply an additional 80% discount factor on the original cost to add an additional subject to compute the cost for 2 grades, 2 subject, and 1 round (2 x 2 x 1). Dividing the resulting cost figure by 2 gives the cost for covering 1 grade and 1 subject (Column V in "tab2.costs.v2"), and dividing the same number by 4 gives the cost of expanding to only 1 additional grade or 1 additional subject (Column W in "tab2.costs.v2"). For MICS, we apply the same calculations with a 20% discount factor.

See detailed costs in sheet "tab2.costs.v2" in the accompanying Excel.

**Box 7. Assessment costs**

If a country participates in a cross-national assessment, the costs consist not only of staff time but also costs related to adapting the test instrument to the national language(s), the training and administration of the test, the per diem for the test administrators, the printing of the tools, the data entry processes, and the data cleaning and consolidation processes, among others. For national assessments, there are additional costs related to test development as well as costs related to dedicated staff.

Internationally administered tests also have additional fixed costs associated with research and development, and a dedicated secretariat in some cases. These costs might be absorbed through fees or through support provided by the host organization or a development partner. There may also be costs for activities beyond measurement of learning, such as policy-data linkages with MICS EAGLE.

With the LDC, per-country costs for organizations currently running regional assessments could fall significantly due to the larger number of countries participating, the greater frequency of assessments, and the provision of standardized global tools, including the item bank, which is a repository of questions and information pertaining to those questions. This would allow these agencies to amortize the fixed costs over a much larger number of countries, resulting in a relative drop in per country costs. However, these efficiency gains may not result in an absolute drop if ambition for the assessment grows in either scope (such as grades or subjects) or technical innovation.

The estimated costs of assessments per round of measurement are summarized in table 2b, including the average fixed costs at the national level per country that vary according to the assessment. The table is built on existing parameters of assessments, such as cycle, features and fees.

*Understanding the Regional Assessment Bundles*

The simulation contains four assessment options or bundles to inform costs for group 3: 1) Policy/Student Linking an existing assessment – typically an NLA. 2) Create a new NLA which includes policy/student linking. 3) Participate in a regional learning assessment. 4) Participate in an international learning assessment

Under the policy/student linking option, countries partner with UIS to set existing assessments against international standards using the Global Proficiency Framework. The NLA and Policy/Student linking bundle assumes costs of starting an assessment from scratch, and requires additional policy/student linking for SDG reporting. Countries can also participate in existing assessments offered in their region, such as SEA-PLM for East Asia and Pacific, or LLECE for Latin America and the Caribbean. Due to regional differences in assessments offered, regions do not have the same options in regional assessments. Costs would then depend on assessment availability in the region. For countries without regional assessments, the international cost is used. Finally, countries can participate in an international assessment (PIRLS, TIMSS, or PISA), where options differ by grades and subjects assessed.

Table 8. Average Costs Per Bundle By Region

| Region | Policy/Student Linking | NLA + Policy/Student Linking | Regional | International |
|--------|------------------------|------------------------------|----------|--------------|
| EAP | $ 324,000 | $ 2,754,000 | $ 833,338 | $ 1,820,568 |
| ECA | $ 324,000 | $ 2,754,000 | | $ 1,820,568 |
| LAC | $ 324,000 | $ 2,754,000 | $ 579,412 | $ 1,820,568 |
| MNA | $ 324,000 | $ 2,754,000 | | $ 1,820,568 |
| NAC | $ 324,000 | $ 2,754,000 | | $ 1,820,568 |
| SAS | $ 324,000 | $ 2,754,000 | | $ 1,820,568 |
| SSA | $ 324,000 | $ 2,754,000 | $ 1,234,453 | $ 1,820,568 |

Note: Blank costs under the regional bundle indicate no regional assessment available. For EAP Regional Assessments, costs reflect SEAPLM. LAC's regional assessment reflect costs of LLECE. For SSA regional assessments, costs reflect the average from PASEC to SACMEQ. For international assessments, cost reflect the average for PISA, PIRLS, and TIMSS.

## Simulating the costs of attaining the goal of 2 x 2 x 1

We simulate how the total cost of doing assessments across 2 grades, 2 subjects, and 1 round in each country changes as the proportion of types of assessments that Group 3 countries (without any assessments used to report on SDG 4.1.1 over the last 7 years or more, or have 1 or more assessments reporting on SDG 4.1.1 that are older than 7 years) do changes. See the Appendix for the different calculations and assumptions driving simulations 1-3, and simulation 4.

There are 2 key input parameters:

- The most recent assessments used to report on SDG 4.1.1. a, b, and c over the last 7 years.
- The corresponding costs of the most recent assessments used to report on SDG 4.1.1 a, b, and c.

Using these 2 inputs parameters, we produce the cost simulations shown below.

- *Scenario 1: Low.* In this scenario, the proportion of Group 3 countries doing policy or student linking, NLA and policy or student linking, regional assessment, and international assessment is 50%, 16.66%, 16.66%, and 16.66% respectively. In this scenario, 50% of Group 3 countries are implementing the relatively lower cost option of policy or student linking, with the other assessment options evenly divided across the remaining 50% of Group 3 countries. Under a "low" scenario, the total cost for implementing assessments across 2 grades, 2 subjects, and 1 round for countries that do not have a recent measure at 2 or more SDG 4.1.1 measurement points is estimated to be around $197 million, reflecting roughly a cost of $1.5 million per country. For IDA/Blend and IBRD countries without a recent measure at 2 or more SDG 4.1.1 measurement points, the total is $89 million and $58 million respectively.
- *Scenario 2: Intermediate.* In this scenario, the proportion of Group 3 countries doing policy or student linking, NLA and policy or student linking, regional assessment, and international assessment is 25%, 25%, 25%, and 25% respectively. In this scenario, the different assessments are evenly divided across Group 3 countries. Under an "intermediate" scenario, the total cost for implementing assessments across 2 grades, 2 subjects, and 1 round for countries that do not

have a recent measure at 2 or more SDG 4.1.1 measurement points is estimated to be around $227 million, reflecting roughly a cost of $1.7 million per country. For IDA and IBRD countries without a recent measure at 2 or more SDG 4.1.1 measurement points, the total is $103 million and $64 million respectively.

- **Scenario 3: High.** In this scenario, the proportion of Group 3 countries doing policy or student linking, NLA and policy or student linking, regional assessment, and international assessment is 16.66%, 50%, 16.66%, and 16.66% respectively. In this scenario, 50% of Group 3 countries are implementing the relatively high cost option of national learning assessment along with policy or student linking, with the other assessment options evenly divided across the remaining 50% of Group 3 countries. Under a "high" scenario, the total cost for implementing assessments across 2 grades, 2 subjects, and 1 round for countries that do not have a recent measure at 2 or more SDG 4.1.1 measurement points is estimated to be around $259 million, reflecting roughly a cost of $2.1 million per country. For IDA and IBRD countries without a recent measure at 2 or more SDG 4.1.1 measurement points, the total is $118 million and $70 million respectively.

For scenarios 1-3, we assign costs for regional assessments depending on the region of the country. For Latin America, we use LLECE. For Sub-Saharan Africa, we take the average of SACMEQ and PASEC. For East Asia and Pacific, we use the cost of SEA-PLM. For regions that do not have a corresponding regional assessment (for example, the Middle East and North Africa, Europe and Central Asia, South Asia) we use the cost of international assessments by taking the average of PIRLS, TIMSS, and PISA. Additionally, users can also choose the numerical proportions of types of assessments (policy or student linking, NLA and policy or student linking, regional assessment, and international assessment) for Group 3 countries.

- **Scenario 4: Country-level Custom.** In this scenario, users can choose two assessments of their choice for each country. Unlike scenarios 1-3, which only allow the user to change the proportion of Group 3 countries doing a particular type of assessment (policy or student linking, NLA with policy or student linking, regional assessment, international assessment), this scenario allows country-level selection of assessments by the users. Therefore, the costs from scenario 4 depend on user choice.

See detailed simulation results in the "output_sheet" in the accompanying Excel.

**Table 9. Assessment costs to achieve 2 x 2 x 2 for countries with no reported learning measures at 2 or more SDG 4.1.1 measurement points in the last 3 years (in millions US$)**

| | Cost Scenarios | | | | | |
| | Low | | Intermediate | | High | |
| | Cost Per Country | Total Cost | Cost Per Country | Total Cost | Cost Per Country | Total Cost |
|---|---|---|---|---|---|---|
| **IDA/Blend** | 3.1 | **179** | 3.6 | **206** | 4.1 | **236** |
| **IBRD** | 3.1 | **116** | 3.5 | **128** | 3.8 | **140** |
| **HIC** | 2.6 | **89** | 3.1 | **108** | 3.6 | **128** |
| **GPE Only\*** | 2.5 | **2.5** | 3.2 | **3.2** | 4.0 | **4.0** |
| **Not Classified** | 3.5 | **7** | 3.9 | **7.7** | 4.3 | **8.6** |
| **All** | 3.1 | **392** | **3.4** | **458** | 3.9 | **518** |

*Note: The three scenarios vary by proportion of different assessment options for group 3 countries. The low cost scenario denotes 50% policy or student linking only, 16.6% NLA + policy or student linking, regional, and international each; the intermediate cost denotes 25% policy or student linking only, NLA + policy or student linking, regional, and international each; The high cost scenario denotes 16.6% policy or student linking only, 50% NLA + policy or student linking, and 16.6% regional, and international each. **GPE Only** represents non-high income GPE countries that are classified into World Bank's lending groups. The only country in this criteria is West Bank and Gaza.*

Note that the figures shown primarily encompass costs of implementing large-scale school assessments. Global public related to classroom assessments and drivers of learning entail additional costs.

## Assumptions

A few key assumptions underlie the cost simulations 1-4:

- We assume that all countries incur the same cost for the same assessment. In practice, this assumption is unlikely to be true as cost for the same assessment may vary by country.
- We assume that regional and international assessments would have the flexibility to be rolled out at the frequency (of grades, subjects, and rounds) to reach the 2 x 2 x 1 goals, if they're currently not doing so.
- We assume that there are economies of scale in expanding to cover more grades and subjects. We assume 80 percent of the total cost to expand an additional grade or subject. For MICS, we assume 20 percent.

Simulations for scenarios 1-3 assume that it is possible that some Group 3 countries have not used an NLA to report on SDG 4.1.1a, b, or c in the last 7 years but could still have an existing NLA (not captured in the data) that could be used for policy linking.

## Costs of multiple rounds

The calculations so far provide costs for implementing assessments across 2 grades, 2 subjects, and 1 round (2 x 2 x 1). However, the LDC's actual goal is for all countries to implement assessments at 2 grades, 2 subjects, and 2 rounds. To implement assessments across 2 rounds instead of 1, we assume that the cost will be double the estimate provided under a 2 x 2 x 1 implementation of assessments. Therefore, under an "intermediate" scenario, the total cost for implementing assessments across 2 grades, 2 subjects, and 2 rounds for countries that do not have a recent learning measure at 2 or more SDG 4.1.1 measurement points is estimated to be around $454 million (227*2), reflecting roughly a cost of $3.4 million (1.7*2) per country. For IDA and IBRD countries without a recent measure at 2 or more SDG 4.1.1 measurement points, the total is $203 million (103*2) and $128 million (64*2) million respectively under an "intermediate" scenario. Under a "low" scenario the total cost for implementing assessments across 2 grades, 2 subjects, and 2 rounds for countries that do not have a recent learning measure at 2 or more SDG 4.1.1 measurement points is estimated to be around $394 million (197*2), reflecting roughly a cost of $3.0 million (1.5*2) per country. The corresponding costs under a "high" scenario is $520 million (260*2), reflecting roughly a cost of $3.9 million (1.9*2) per country.

The cost calculations shown above demonstrate that it is fiscally feasible for the world to end the learning data crisis. It would take a little more than 500 million dollars to ensure that all countries have learning assessments for 2 grades, 2 subjects, and 2 rounds. To ensure that countries can cost effectively attain the goal of 2 grades, 2 subjects, and 2 rounds (2 x 2 x 2) to assess student learning, the LDC builds flexibility into its costing model, so that countries may choose assessments most suitable to their context and constraints. For reporting countries that have assessed one grade or one subject over the last 7 years, and countries that have not reported any measurements in the last 7 years, there are low, intermediate, and high cost assessment options to attain the goal. This model also allows reporting countries (over the last 7 years) to build on existing assessments and mitigate costs rather than start from scratch.

While having learning data alone does not automatically solve the learning crisis, without such data the mission of ending learning poverty is impossible. Only with multiple rounds of temporally comparable data can we identify what's working to improve learning, identify and rectify gaps and needs, and support the efforts of teachers and parents to recover learning losses caused by Covid-19. Such data is invaluable to direct our response to the pandemic, and address the learning crisis that predated Covid-19.

Note that the figures shown primarily encompass costs of implementing large-scale school assessments under Pillar 1. Costs for facilitating data use are additional. Similarly, costs for global public goods and tools related to classroom assessments and drivers of learning entail additional costs. Further details about the costing exercise can be found in the accompanying costing note.

*Pillar 2: Measuring the Drivers of Learning: Estimated Cost*

**Table 10. Estimates Costs for Pillar 2**

| Tool | Cost |
| --- | --- |
| School Surveys with Classroom Observations | ~$100K for 100 classrooms |
| School Surveys | ~$1300K to 300K per country |
| Household Surveys | ~$1.5 million for a nationally representative sample (and adding a learning modules to an existing household survey 500K) |
| Administrative Records | ? |

The estimated cost to develop the pilots and methodology on how to integrate six learning modules into national household survey will be between $500k-1m per country, depending on country choice and local costs. For instance, for Uganda the additional cost would be in the ballpark of $700-750k, including analytical support and report writing, data preparation, documentation, and dissemination, etc.

Notice that this cost builds on the existing investment that the country is already making in running their household surveys. We do not expect to cover in the pilot the full cost of running a multitopic household survey, which on average is estimated to cost approximately 2 million dollars.

We also estimate that the pilot costs will be higher than the scale-up phase, since at this stage we will also be covering the development of the tools and creation and documentation of protocols that will need to be used at the scale-up phase.

This work will also leverage on other existing indicatives such as the UNICEF improvement of the Foundational Learning Module (FLS), which measures 4.1.1.(a). Funded by Bill & Melinda Gates Foundation, UNICEF is going to revise the existing FLS to collect more internationally comparable learning outcomes data across countries. In addition, the new FLS aims to measure 4.1.1.(b) building on existing tools and accumulated experience in collecting learning data at households. It's expected that the project will identify cost-effective approaches to promote student learning, facilitate the understanding of the barriers girls face in participating at school, and accelerate results towards achieving SDG4 on quality education for all children. In addition, UNICEF is piloting linking household survey data to administrative data in order to capture both supply-side (student's individual and household characteristics) and demand-side factors (e.g., teacher qualification, school infrastructure, school budget) to understand the impacts on these factors on learning outcomes. These innovative data approaches are a great complement to school-based full-fledged learning assessments by shedding light on unanswered questions to address the learning crisis.

EMIS: Most countries have some type of EMIS in place, varying in their degree of sophistication in terms of data collection, management, and use. Cost of improving EMIS will vary depending on the starting point and the level of ambition of the goal.

Cost of implementing school surveys varies immensely. XXX

While the typical cost of a school survey is $300,000 per country, GEPD has also focused efforts to reduce average costs to $130,000. This cost covers the 1) survey firm, 2) local consultant(s), 3) and necessary travel for trainers. Excluded from these estimates is the staff time required to analyze the data; however, data processing and calculation of indicators have been largely automated, so that little to no staff time is needed.

Survey modules would be made available for use free of charge. The cost of administering the survey or of possible advisory services that may accompany conducting such a survey would need to be covered.

Based on MICS estimates, the average cost to implement a nationally representative household survey is approximately 1 million.

## *Pillar 3: Understand Classroom Practices: Estimated Cost*

- Integrate tools and approaches into teacher, curriculum, and school planning processes; and,
- Ensure strong linkages to national assessments and data systems.

**Table 11. Estimated cost**

| | Tool | Target | Subjects | Administration | Agency | Cost per country |
|---|---|---|---|---|---|---|
| Classroom assessments | Formative Assessment (ASER-based) | Grades 3-6 | Math Reading | Teachers | UNICEF/ Pratham | |
| | Formative Assessment (ASER-based) | Teachers | Math Reading | Teacher Trainers | UNICEF/ Pratham | $390,000 |
| | ECARO Teachers Guide | Teachers | | Teachers | *UNICEF ECARO* | |
| School Classroom Research and | Early Learning Assessment | Pre-primary | | Teachers | | $150,000 |
| | Classroom observation | | | | | $200,000 |
| | Data Must Speak | Whole school | NA | All | | $300,000 |
| | Time to Teach | Whole school | All | All | | $140,000 |

Collecting and using data the ***classroom assessments for learning*** will incur high initial start-up costs related to the building of teacher capacity. It is important that a critical mass of personnel within a school are supported in accessing training and ongoing support to develop deep understanding and can easily use the tools and approaches. It is also essential that the initial investment be supported with subsequent refresher training and integration into routine capacity building and training regimes. Therefore, costing takes into consideration support for initial training including coaches and master trainers, support for integration into teacher training and development programmes, tool modification, and periodic support for two years.

# 7. Operational Arrangements

The LDC is a shared vision among the three agencies (UNESCO, WB, and UNESCO) and development partners who join them to ensure that every low- and middle-income country will measure learning in at least two subjects, in two grades, over at least two rounds within a five-year period. This data will be complemented through data on drivers of learning and classroom practices. To operationalize this vision, the LDC lays out a set of shared principles, processes, and accountabilities to guide technical assistance and funding (through loans and/or grants) from development partners and donors on learning measurement over the next five years. At the end of this period, the effectiveness of the Compact should be evaluated, and its future and scope reassessed.

The compact has two main components: activities at the country level, and global and regional public goods. The country component demands coordination in funding and, very likely, new investments, including domestic mobilization of resources. For both components, the program will rely on the shared principles, processes, and accountabilities to coordinate individual donor contributions.

The funding pledges and disbursements, as well as technical pledges, would be brokered, coordinated, and tracked to ensure accountability and efficient use of the resources currently and in the future. Most funding will be channeled directly to countries. The LDC Technical Secretariat will coordinate the existing global and regional public goods that may be provided to countries to support the processes required to the deliver on the LDC's shared vision.

Figure 12 illustrates how different organizations might use different instruments to finance the implementation of the LDC both at the country level and in the production of global and regional public goods. At the country level, in addition to domestic resource mobilization, each agency will use instruments at their disposal, such as grants or lending, to support country efforts, with the clear understanding that in some cases donors will be encouraged by the LDC to directly fund countries. Exactly what is funded, for example, data collection, capacity building, assessment design, etc. will depend on the needs and demands of the country. For the production of public goods, agencies will use existing and potential new trust fund resources.

**Figure 12. Funding mechanisms for the Learning Development Compact**
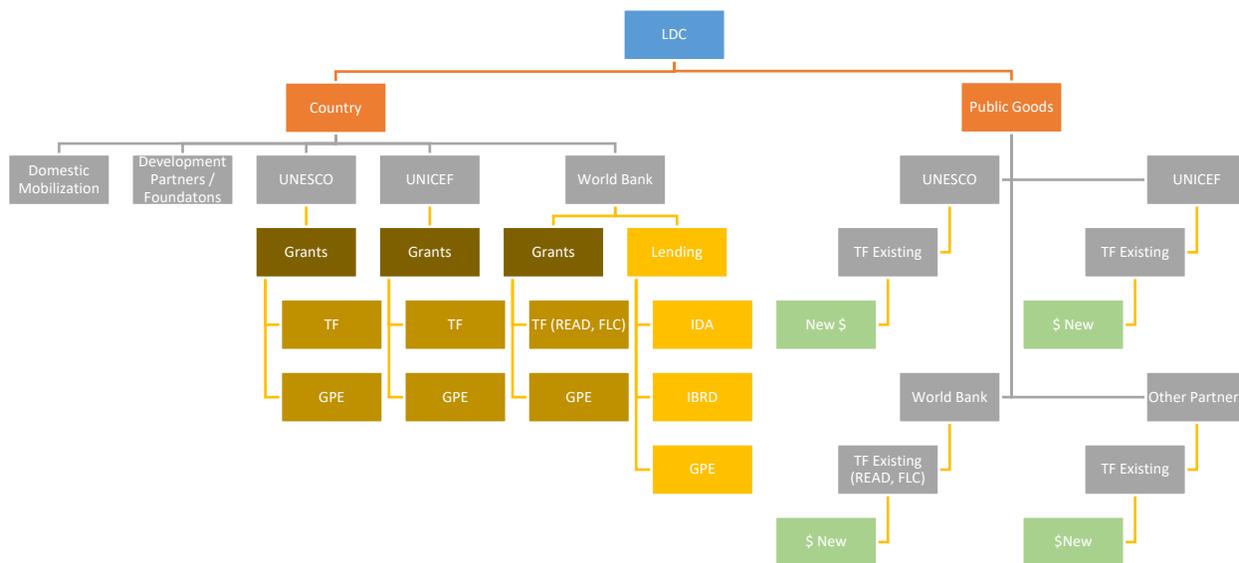


### A Phased Approach to Working with Countries

The LDC will help countries prioritize measurement points for assessing learning. Ideally, countries would measure learning at all three SDG 4.1.1 measurement points: grades 2/3, end of primary, and end of lower secondary (Pillar 1). This data can be complemented with data on drivers of learning and classroom practices. But countries face limited budgets and inevitably need to make choices about where to focus scarce resources for measuring learning. The LDC will need to make decisions on two fronts:

1. What partner countries to prioritize for supporting under the Compact
2. What pillars of the LDC to prioritize, which may differ depending on the country
3. Which SDG 4.1.1 measurement points to prioritize for support under the LDC

**On the question of what countries to prioritize,** the immediate focus of the LDC for country-level prioritization[22] will be to support countries with the greatest assessment gaps (such as, Group 3 countries) and the lowest income levels, before gradually expanding to all other countries in a phased way (see Figure 13). Other factors such as demand from countries will also play a critical role in country prioritization. The phased approach is being developed based on an initial draft breakdown of countries using key variables, such as assessment gaps and country income classification. Out of 145 low-, lower-middle, and upper-middle income countries, we have identified 49 countries that have data for at least two recent SDG 4.1.1 measurement points. Of countries that do not have at least two recent learning measures, 59 are World Bank IDA/Blend/ GPE eligible countries and the remaining 37 are eligible for World Bank's IBRD funding. As mentioned, the goal of the LDC is to move countries toward a 2 x 2 x 2 structure to implement learning assessments for at least 2 grades,[23] in at least 2 subjects, for at least two rounds by 2030. Under a phased approach, the LDC will start with Group 3[24] countries in Phase I (see Figure 13)—countries that have no recent[25] learning measurement at any of the three SDG measurement points, with a particular focus on the low- and lower middle-income. The Compact will gradually expand coverage to include Group 2b, countries that have reported only on one SDG measurement point over the last 7 years, again with a focus on the low- and lower middle-income countries. This phase may also

include expanding coverage to Group 2a countries that have reported on two or more SDG measurement points but not recently (Group 2a). Groups 2a, 2b and 3 include countries that are currently falling short of the goal to measure learning for at least 2 grades, and the LDC can help them move toward that goal. Eventually in Phase 3, the Compact will expand to cover all remaining countries, including those in Group 1 as needed. A phased approach will enable the initiative to gradually expand support and funding in different phases, while ensuring that lessons from Phase I inform future phases of the Compact.

**Figure 13. Number of countries and percentage of school age population covered in each group**
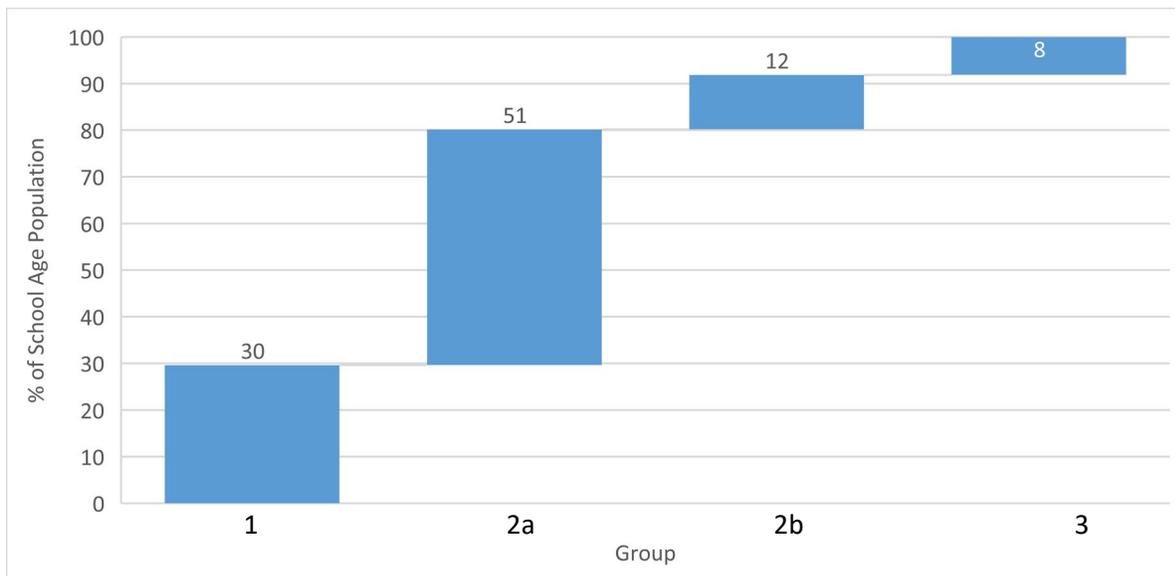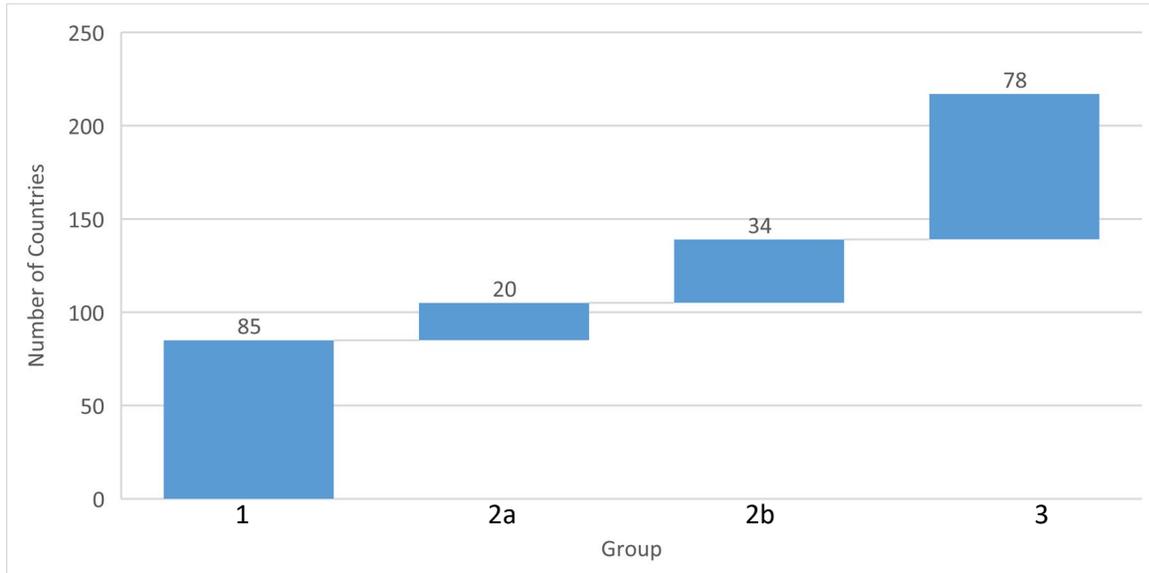




*Note: Group 1 countries have reported on 2 or more SDG measurement points through assessments that took place in the last 3 years. Group 2a countries have reported on 2 or more SDG measurement points through assessments that took place more than 3 years ago but less than 7 years ago. Group 2b countries have only reported on 1 SDG measurement point through assessments that took place over the last 7 years. Group 3 countries have reported on either no SDG measurement points or have reported on one or more SDG measurement point through an assessment that is older than 7 years. Percentage of primary age population uses population aged 5-16.*

***On the question of what pillars to prioritize in countries,*** the LDC will prioritize among pillars (large-scale assessments, data on drivers of learning, classroom-based assessments) based on key principles mentioned earlier, including building on what already exists (Principle 1), flexibility based on country needs (Principle 2), country demands (Principles 3), and relevance of data collected to country-level decision-making (Principle 4). Therefore, the recommendations are likely to vary depending on the country. For example, if a country has a national assessment in place, it may prioritize participating in a policy-linking exercise to use the national assessment to report on SDG 4.1.1 benchmarks. If no national assessment exists that may be used for a policy-linking exercise, a country may prioritize participating in an existing regional or international assessment. Similarly, countries that have existing household surveys can opt to add MICS foundational learning modules to enable SDG reporting. Furthermore, countries may make investments to complement such data with data on drivers of learning and classroom practices to make informed decisions, depending on country needs, demands, and utility for decision-making.

***On the question of what SDG 4.1.1 measurement points to prioritize in partner countries for Pillar 1,*** the LDC builds on the principles lister earlier, building on what exists and policy relevance. More concretely, some of the guiding question can be:

- **Building on what exists:** Countries may make investment decisions based on where the assessment gaps lie. For example, if there is no recent measure of learning at early or end of primary, countries could consider investing in learning assessments at those earlier levels. To ensure temporal comparability, countries may prioritize continuing measuring learning through (temporally comparable) assessments that have already been recently implemented (Principle 1). While the LDC will encourage countries follow key principles when choosing assessments, such as building on what exists, and relevance of data to decision-making (Principle 1 and 4), these decisions will ultimately depend on country context, needs, and demands (Principle 2 and 3).
- **Policy relevance:** Countries may also prioritize measuring learning based on the grade levels where most of the enrolled populations are. For example, countries with limited enrollment at the secondary level may choose to invest resources to measure learning at the primary level, where the bulk of the student population might be enrolled.

**Group 3: Countries with no recent assessment at any 3 SDG measurement points.** 51 IDA/Blend/GPE countries (of which 15 are low-income and 12 are lower-middle income) [26][27][28] have not reported on SDG a, b, *or* c measurement points in the last three years. Countries that have no recent learning assessment at any of the three SDG measurement points may prioritize measurement at grades 2/3 and end of primary, as their enrollment in lower-secondary is likely to be low. It is important to identify learning deficiencies early in the schooling cycle so that appropriate remediation mechanisms can be implemented in a timely manner. Since learning is a cumulative process, fixing learning deficiencies at earlier education levels can have positive impacts on learning in later grades. In alignment with the 2 x 2 x 2 vision, such countries should commit to doing at least two rounds of data collection for each learning assessment (over 5 years) by 2030.

If such countries have high secondary enrollment, they may choose to prioritize one measurement point at either early primary or end of primary, and the second measurement point at lower secondary. Among countries that do not have a recent measure of SDG 4.1.1a, b, and c and have available information on secondary enrollment, no low-income countries, 2 lower-middle income countries (Sri Lanka and Uzbekistan), and 1 (Uzbekistan) IDA/Blend/GPE country have secondary enrollments greater or equal to 90 percent. [29][30]

Based on the foregoing discussion, countries with no recent measurement at SDG 4.1.1a, b and c may consider the following three packages:

- Assessments to measure learning at SDG 4.1.1 a and b (for most countries)
- Assessments to measure learning at SDG 4.1.1 b and c (for the few countries with high secondary enrollment)
- Assessments to measure learning at SDG 4.1.1 a and c (for the few countries with high secondary enrollment)

**Group 2b and 2a: Countries with a recent learning measurement at one SDG measurement point only (Group 2b), or measurement at 2 or more measurement points that are from more than 3 years and less than 7 years ago (Group 2a).**

**Group 2b.** Only 1 low-income country and 6 lower middle-income countries have *only* reported on SDG 4.1.1a in the last three years. No low-income countries and 7 lower-middle income countries have *only* reported on SDG 4.1.1b in the last three years. No low-income countries and 2 lower-middle income countries have *only* reported on SDG 4.1.1c in the last three years. Among IDA/Blend/GPE countries, 6 countries have a recent measure at SDG 4.1.1a only, 5 have a recent measure at SDG 4.1.1b only, while none have a recent measure at SDG 4.1.1c only.

Among countries that have a recent measure at one SDG measurement point only, and have available information on secondary enrollment, no low-income countries, lower middle-income countries, or IDA/Blend/GPE countries have secondary enrollment greater or equal to 90 percent. Given the low prevailing secondary enrollment across Group 2b countries, such countries may prioritize measuring learning at early primary and end of primary. In alignment with the 2x2x2 vision, they should commit to doing at least two rounds of data collection for each learning assessment (over 5 years). Broadly speaking, countries with a recent assessment at only one SDG measurement point may consider the following packages:

- Assessments to measure learning at SDG 4.1.1 a and b.

**Group 2a.** Only 9 low-income countries and 27 lower middle-income countries have reported on two or more SDG measurement points through assessments that took place more than 3 years and less than 7 years ago. Among countries that have reported on two or more SDG measurement points through assessments that took place more than 3 years and less than 7 years ago, and have available information on secondary enrollment, no low-income countries, 2 lower middle-income countries, and no IDA/Blend/GPE countries have secondary enrollment greater or equal to 90 percent (see Appendix). In alignment with the 2x2x2 vision, they should commit to doing at least two rounds of data collection for each learning assessment (over 5 years). Broadly speaking, countries with assessment data at two SDG measurement points from more than 3 years and less than 7 years ago may consider the following packages:

- Assessments to measure learning at SDG 4.1.1 a and b (for most countries)
- Assessments to measure learning at SDG 4.1.1 b and c (for the few countries with high secondary enrollment)
  Assessments to measure learning at SDG 4.1.1 a and c (for the few countries with high secondary enrollment)

**Group 1** includes countries that have a recent measure of learning at 2 or more points at SDG measurement points. Currently, 16 IDA/Blend/GPE countries and 33 IBRD countries have a recent learning measure at 2 or more SDG measurement points. Similarly, 8 low-income countries, 12 lower middle-

income countries, and 27 upper middle-income countries fall in this category. Following some of the principles listed above, such as building on what exist and country demand, the LDC approach encourages these countries to continue to measure learning with the assessment programs they have already been using and benefit from global public goods that the initiative produces.

Note that while the information above is meant to serve as principles to gauge existing capacities and needs. Each country's unique context and political economy for implementing and using assessments will determine investment needs. See appendix tables A1 and A2 for a full mapping of country SDG reporting status (over the last 3, 5, and 7 years) and secondary enrollment (latest available figure since 2013) by country income and lending classifications.

Ideally all recipient countries would be expected to sign up to a set of principles outlining how they would work with development partners, and the Compact more broadly, to maximize impact. These principles will set forth ways of working, regular review processes, and communication frameworks. They will also outline a co-financing and contributions framework, to encourage countries to gradually take greater ownership of efforts to improve learning data, as evidenced by domestic financing and policy initiatives. These principles will be devised in collaboration with countries and development partners in the coming months to ensure relevance and feasibility.

## 8.  Next Steps, Milestones, and Results Indicators

The LDC, UNESCO, UNICEF, and the World Bank are conducting a number of consultations with multiple stakeholders to introduce the vision, as well as shared principles, processes and accountabilities underlying the LDC. The LDC has and will leverage major convenings such as the Global Education Summit, United National General Assembly (UNGA), World Bank Annual Meetings and UNESCO General Assembly to organize side events to present, discuss, and solicit feedback on technical elements of the LDC. We also plan to mobilize additional resources at the international level to allow the successful implementation of the vision, including both the country engagement as well as the accompanying production of public goods. See Figure 14 for a visual illustration of our LDC timeline and critical milestones.

Under the current proposal, the plan is to officially launch the LDC in late 2021 or early 2022, organize the first meeting of the steering committee, endorse the plan, and move to country implementation. The production of public goods has already started and will continue, and the plan is to increase both the coordination and velocity of producing them. To address the increased urgency of the learning crisis due to COVID-19, early implementation will focus efforts to measure learning losses.

As discussed in the accountabilities section of this document, the proposal is to have bi-annual meetings of the steering committee for stocktaking and discussing the progress report of the LDC. We are also proposing two process evaluations of the LDC, which will assess both the intended and unintended consequences of the LDC.

**Figure 14. Proposed Timeline and Milestones**
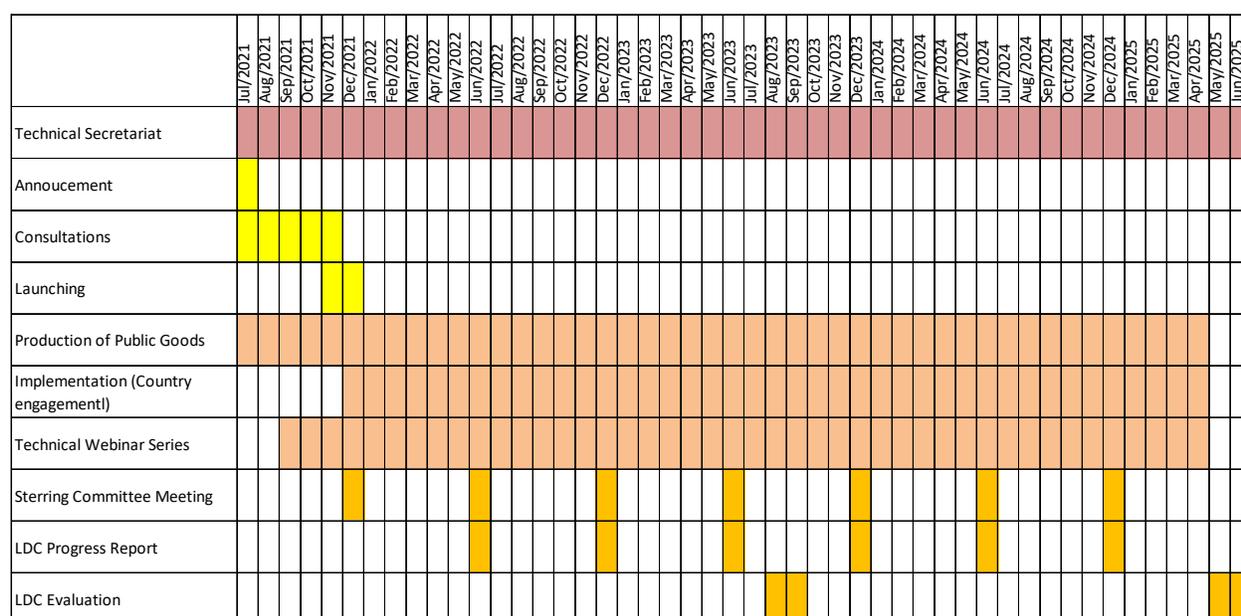


The LDC's success will be defined in terms of the availability of temporally comparable learning data, and progress can be monitored based on a few selected indicators (Table 12), building on baseline values of learning assessment availability in 2018–2020 period. There are two main groups of indicators, those reflecting the countries currently on track (implementing learning assessments with at least the 2 x 2 x 2 frequency) and countries that are off-track.

For the countries on-track, currently 16 IDA/Blend/GPE and 33 IBRD, the LDC success will be determined based on the fact that all those countries continue to report measures of learning with similar quality (indicator 1). The remaining set of indicators reflect the objectives of the LDC for the off-track countries, and reflect three main steps: countries should plan how to report temporally comparable measures of learning, report measures of learning, and ensure that these measures of learning are temporally comparable. This last indicator should be reported cumulatively, while the others should focus on the number of countries delivering on those results every year.

**Table 12. Results indicators, baseline values, and targets through 2025**

| | Baseline | Jun-22 | Jun-23 | Jun-24 | Jun-25 |
|---|---|---|---|---|---|
| # IDA/Blend/GPE countries currently monitoring learning | 16 | | 16 | | |
| # of NEW IDA/Blend/GPE countries planning learning assessments | | 20 | 20 | 19 | |
| # of NEW IDA/Blend/GPE countries monitoring and reporting on learning | | | 20 | 20 | 19 |
| # Number of IDA/Blend/GPE temporally comparable measures of learning | 16 | | 36 | 56 | 75 |
| # IBRD countries currently monitoring learning | 33 | | | | |
| # of NEW IBRD countries planning learning assessments | | 13 | 12 | 12 | |
| # of NEW IBRD countries monitoring and reporting learning | | | 13 | 12 | 12 |
| # Number of IBRD temporally comparable measures of learning | 33 | | 46 | 58 | 70 |

The baseline shows countries currently on track to monitor learning, while additional years suggest the number of participating countries per year to achieve learning measurements worldwide by 2025.

# 9. References

Abdul-Hamid, Husein. 2014." What Matters Most for Education Management Information Systems: A Framework Paper." World Bank, Washington, DC. http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/Background/EMIS/Framework_SABER-EMIS.pdf

All Children Learning. "All Children Learning – Assessment Platform." https://allchildrenlearning.org/

Azevedo, João Pedro. 2020. Learning Poverty: Measures and Simulations. Policy Research Working Paper; No. 9446. World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/34654

Azevedo, João Pedro, Diana Goldemberg, Silvia Montoya, Reema Nayar, Halsey F. Rogers, Jaime Saavedra, and Brian William Stacy. 2021a."Will Every Child Be Able to Read by 2030? Defining Learning Poverty and Mapping the Dimensions of the Challenge," Policy Research Working Paper Series 9588, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/35300

Azevedo, João Pedro, Amer Hasan, Diana Goldemberg, Koen Geven, Syedah Aroob Iqbal. 2021b. Simulating the Potential Impacts of COVID-19 School Closures on Schooling and Learning Outcomes: A Set of Global Estimates, The World Bank Research Observer, Volume 36, Issue 1, February, Pages 1–40, https://doi.org/10.1093/wbro/lkab003

Bruns, Barbara, Maryam Akmal, and Nancy Birdsall. 2019. "The Political Economy of Testing in Latin America and Sub-Saharan Africa." CGD Working Paper, World Bank, Washington, DC. https://www.cgdev.org/sites/default/files/political-economy-testing-latin-america-and-sub-saharan-africa-dec2019.pdf

Clarke, Marguerite. 2015. *Are citizen-led assessments raising learning levels?* Washington, DC: World Bank. https://blogs.worldbank.org/education/are-citizen-led-assessments-raising-learning-levels

Clarke, Marguerite and Diego Luna-Bazaldua. 2021. "Primer on Large-Scale Assessments of Educational Achievement. World Bank, Washington, DC. https://openknowledge.worldbank.org/bitstream/handle/10986/35494/9781464816598.pdf?sequence=4.

Cofré, Carlos Cayumán. "UNESCO´s Large Scale Student Assessment Program in LAC." http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/05/GAML6-Session5-LLECE.pdf

Hoyos, Rafael de, Vicente A. Garcia-Moreno, and Harry Anthony Patrinos. 2017. "The Impact of an Accountability Intervention with Diagnostic Feedback: Evidence from Mexico." *Economics of Education Review* 58: 123-140. https://doi.org/10.1016/j.econedurev.2017.03.007

Hoyos, Rafael de, Alejandro Ganimian, Peter Anthony. 2017. "Teaching with the Test: Experimental Evidence on Diagnostic Feedback and Capacity Building for Public Schools in Argentina." Policy Research Working Paper Series 8261, World Bank, Washington, DC. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/488751511886071513/teaching-with-the-test-experimental-evidence-on-diagnostic-feedback-and-capacity-building-for-public-schools-in-argentina

ICCS. "International Civic and Citizenship Education Study." https://www.iea.nl/studies/iea/iccs

Jordan, Rachel. 2019. "Science of Teaching: Classroom Level." RTI International. Research Triangle Park, NC. https://scienceofteaching.s3.eu-west-3.amazonaws.com/index.html#/lessons/TthxFqWCoRmHK1fVvm1fhWQdpgGvS_fh

OECD. "OECD Survey on Social and Emotional Skills." https://www.oecd.org/education/ceri/social-emotional-skills-study/

PISA. "PISA 2018 Global Competence." https://www.oecd.org/pisa/innovation/global-competence/

Oubahi, Yasmine. 2021. "Measuring Life Skills: Towards a Transformative Vision of Education." https://www.unicef.org/mena/stories/measuring-life-skills-towards-transformative-vision-education

Results for Development Institute (R4D). 2015. "Bringing Learning to Light: The Role of Citizen-Led Assessments in Shifting the Education Agenda." Results for Development. Washington, DC: http://r4d.org/resources/bringing-learning-light-role-citizen-led-assessments-shifting-education-agenda/

Sandoval-Hernández, Andrés, Maria Magdalena, and Daniel Miranda. "Measurement Strategy for SDG Global Indicator 4.7.1 and Thematic Indicators 4.7.4 and 4.7.5 using International Large Scale Assessments in Education." http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/08/GAML6-REF-9-measurement-strategy-for-4.7.1-4.7.4-4.7.5.pdf

SEA-PLM. 2020. "SEA-PLM 2019 Main Regional Report Summary." https://www.unicef.org/eap/media/7631/file

SEA-PLM. 2020. "SEA-PLM 2019 Main Regional Report." https://www.unicef.org/eap/media/7356/file/SEA-PLM%202019%20Main%20Regional%20Report.pdf

UNESCO Institute for Statistics Technical Cooperation Group. 2020. "Policy Linking". UNESCO Institute of Statistics. Montreal, Canada. http://tcg.uis.unesco.org/policy-linking/

UNESCO Institute of Statistics, US Agency for International Development, UK Department for International Development, Australian Council for Educational Research, and the Bill and Melinda Gates Foundation. "Global Proficiency Framework for Reading and Mathematics." UNESCO UIS. http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/05/Global-Proficiency-Framework-18Oct2019_KD.pdf

UNESCO. "Indicator 4.7.5." http://gaml.uis.unesco.org/indicator-4-7-5/

UNICEF Innocenti. 2021. *Time to Teach.* Florence, Italy: UNICEF. https://www.unicef-irc.org/research/time-to-teach/

UNICEF. 2019. *Every Child Learns: UNICEF Education Strategy 2019 – 2030.* New York, NY: UNICEF. https://www.unicef.org/media/63056/file/EdStrategy-2019-2030-CountrySolution-WCAR.pdf

Wagner, Daniel A, Andrew Babson, and Katie M. Murphy. 2011. "How Much is Learning Measurement Worth? Assessment Costs in Low-Income Countries" *Current Issues in Comparative Education* 14: 3-21. https://repository.upenn.edu/literacyorg_articles/2

World Bank. 2019. *Ending Learning Poverty: What Will It Take?* Washington, DC: World Bank. http://hdl.handle.net/10986/32553.

World Bank. 2021. *World Development Report: Data for Better Lives*. Washington, DC: World Bank. https://www.worldbank.org/en/publication/wdr2021.

# 10. Annex

**Table A1. SDG reporting coverage by country income groupings**

| | | Reporting on SDG 4.1.1 over last 3 years (2018 or later) | | | Reporting on SDG 4.1.1 over last 5 years (2016 or later) | | | Reporting on SDG 4.1.1 over last 7 years (2014 or later) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Lower secondary enrolled >= 90% | Lower secondary enrolled < 90% | No enrollment data | Lower secondary enrolled >= 90% (latest) | Lower secondary enrolled < 90% | No enrollment data | Lower secondary enrolled >= 90% (latest) | Lower secondary enrolled < 90% | No enrollment data |
| **Global** | Don't report on SDG 4.1.1a, b **&** c | 5 | 47 | 44 | 4 | 39 | 40 | 4 | 38 | 36 |
| | SDG 4.1.1a | - | 6 | 5 | - | 10 | 7 | - | 10 | 6 |
| | SDG 4.1.1b | - | 6 | 3 | - | 4 | 4 | - | 3 | - |
| | SDG 4.1.1c | 5 | 11 | - | 4 | 10 | - | 3 | 10 | 2 |
| | SDG 4.1.1a & b | 1 | 25 | 6 | 1 | 27 | 6 | 1 | 25 | 7 |
| | SDG 4.1.1b & c | 30 | 13 | 4 | 32 | 16 | 4 | 33 | 18 | 8 |
| | SDG 4.1.1a & c | 3 | 1 | - | 3 | 1 | 1 | 3 | 1 | 2 |
| | SDG 4.1.1a, b **&** c | 2 | 1 | - | 2 | 3 | - | 2 | 5 | 1 |
| **Low-income** | Don't report on SDG 4.1.1a, b **&** c | - | 13 | 7 | - | 10 | 6 | - | 10 | 5 |
| | SDG 4.1.1a | - | - | 1 | - | 2 | 2 | - | 2 | 2 |
| | SDG 4.1.1b | - | - | - | - | - | | - | - | - |
| | SDG 4.1.1c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a & b | - | 1 | 7 | - | 7 | 1 | - | 7 | 1 |
| | SDG 4.1.1b & c | - | - | - | - | 1 | - | - | 1 | - |
| | SDG 4.1.1a & c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a, b **&** c | - | - | - | - | - | - | - | - | - |
| **Lower-middle income** | Don't report on SDG 4.1.1a, b **&** c | 2 | 15 | 6 | 1 | 11 | 5 | 1 | 10 | 2 |
| | SDG 4.1.1a | - | 3 | 3 | - | 4 | 4 | - | 4 | 3 |
| | SDG 4.1.1b | - | 5 | 2 | - | 3 | 2 | - | 2 | - |
| | SDG 4.1.1c | - | 2 | - | - | 2 | 1 | - | 3 | 2 |
| | SDG 4.1.1a & b | - | 7 | 3 | - | 9 | 3 | - | 7 | 4 |
| | SDG 4.1.1b & c | - | 2 | - | - | 3 | - | 2 | 4 | 1 |
| | SDG 4.1.1a & c | - | - | - | - | - | - | - | - | 1 |
| | SDG 4.1.1a, b **&** c | - | - | - | - | 2 | - | - | 4 | - |
| **Upper-middle income** | Don't report on SDG 4.1.1a, b **&** c | - | 10 | 11 | - | 9 | 10 | - | 9 | 10 |
| | SDG 4.1.1a | - | 3 | 1 | - | 4 | 1 | - | 4 | 1 |
| | SDG 4.1.1b | - | 1 | 1 | - | 1 | 1 | - | 1 | - |
| | SDG 4.1.1c | - | 3 | - | - | 3 | - | - | 2 | - |
| | SDG 4.1.1a & b | 1 | 9 | 2 | 1 | 9 | 2 | 1 | 9 | 2 |
| | SDG 4.1.1b & c | 4 | 6 | 3 | 4 | 6 | 3 | 4 | 7 | 4 |
| | SDG 4.1.1a & c | 1 | - | - | 1 | - | 1 | 1 | - | 1 |
| | SDG 4.1.1a, b **&** c | - | 1 | - | - | 1 | - | - | 1 | - |
| **High-income** | Don't report on SDG 4.1.1a, b **&** c | 3 | 9 | 20 | 3 | 9 | 19 | 3 | 9 | 19 |
| | SDG 4.1.1a | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1b | - | - | - | - | - | 1 | - | - | - |
| | SDG 4.1.1c | 5 | 6 | - | 3 | 5 | - | 3 | 5 | - |
| | SDG 4.1.1a & b | - | 2 | - | 0 | 2 | - | - | 2 | - |
| | SDG 4.1.1b & c | 26 | 5 | 1 | 28 | 6 | 1 | 28 | 6 | 2 |
| | SDG 4.1.1a & c | 2 | 1 | - | 2 | 1 | - | 2 | 1 | - |
| | SDG 4.1.1a, b **&** c | 2 | - | - | 2 | - | - | 2 | - | - |

Note: SDG 4.1.1a reporting for high-income countries tends to rely on PIRLS grade 4 whereas other countries use assessments done at grades 2/3.

# Table A2. SDG reporting coverage by lending status

| | | Reporting on SDG 4.1.1 over last 3 years (2018 or later) | | | Reporting on SDG 4.1.1 over last 5 years (2016 or later) | | | Reporting on SDG 4.1.1 over last 7 years (2014 or later) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Lower secondary enrolled >= 90% | Lower secondary enrolled < 90% | No enrollment data | Lower secondary enrolled >= 90% | Lower secondary enrolled < 90% | No enrollment data | Lower secondary enrolled >= 90% | Lower secondary enrolled < 90% | No enrollment data |
| **IDA/Blend/GPE only** | Don't report on SDG 4.1.1a, b & c | 1 | 33 | 14 | 1 | 27 | 13 | 1 | 26 | 10 |
| | SDG 4.1.1a | - | 4 | 2 | - | 3 | 7 | - | 7 | 3 |
| | SDG 4.1.1b | - | 4 | 1 | - | 2 | 1 | - | 2 | - |
| | SDG 4.1.1c | - | - | - | - | - | - | - | 1 | 1 |
| | SDG 4.1.1a & b | - | 12 | 4 | - | 14 | 4 | - | 12 | 5 |
| | SDG 4.1.1b & c | - | - | - | - | 2 | - | - | 2 | 1 |
| | SDG 4.1.1a & c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a, b & c | - | - | - | - | 1 | - | - | 3 | - |
| **IBRD** | Don't report on SDG 4.1.1a, b & c | 2 | 9 | 10 | 1 | 7 | 8 | 1 | 7 | 7 |
| | SDG 4.1.1a | - | 3 | 2 | - | 3 | 3 | - | 3 | 2 |
| | SDG 4.1.1b | - | 2 | 2 | - | 3 | 2 | - | 1 | - |
| | SDG 4.1.1c | - | 7 | - | 1 | 7 | - | - | 6 | 1 |
| | SDG 4.1.1a & b | 1 | 12 | 2 | 1 | 12 | 2 | 1 | 12 | 2 |
| | SDG 4.1.1b & c | 6 | 8 | 3 | 6 | 8 | 3 | 7 | 10 | 6 |
| | SDG 4.1.1a & c | - | - | 1 | 1 | - | 1 | 2 | - | 1 |
| | SDG 4.1.1a, b & c | - | - | - | - | 1 | - | - | 1 | - |
| **High-income countries** | Don't report on SDG 4.1.1a, b & c | 2 | 5 | 18 | 2 | 5 | 18 | 2 | 5 | 18 |
| | SDG 4.1.1a | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1c | 4 | 5 | - | 3 | 3 | - | 3 | 3 | - |
| | SDG 4.1.1a & b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1b & c | 24 | 5 | 1 | 26 | 6 | 1 | 26 | 6 | 1 |
| | SDG 4.1.1a & c | 2 | 1 | - | 2 | 1 | - | 2 | 1 | - |
| | SDG 4.1.1a, b & c | 2 | - | - | 2 | - | - | 2 | - | - |
| **Not classified** | Don't report on SDG 4.1.1a, b & c | - | - | 2 | - | - | 1 | - | - | 1 |
| | SDG 4.1.1a | - | - | - | - | - | 1 | - | - | 1 |
| | SDG 4.1.1b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a & b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1b & c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a & c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a, b & c | - | - | 1 | - | - | 1 | - | - | 1 |

Note: SDG 4.1.1a reporting for high-income countries tends to rely on PIRLS grade 4 whereas other countries use assessments done at grades 2/3.

**Table A3. SDG reporting coverage by Region**

| | | Reporting on SDG 4.1.1 over last 3 years (2018 or later) | | | Reporting on SDG 4.1.1 over last 5 years (2016 or later) | | | Reporting on SDG 4.1.1 over last 7 years (2014 or later) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Lower secondary enrolled >= 90% | Lower secondary enrolled < 90% | No enrollment data | Lower secondary enrolled >= 90% | Lower secondary enrolled < 90% | No enrollment data | Lower secondary enrolled >= 90% | Lower secondary enrolled < 90% | No enrollment data |
| EAP | Don't report on SDG 4.1.1a, b & c | - | 7 | 12 | - | 7 | 10 | - | 7 | 9 |
| | SDG 4.1.1a | - | 1 | 2 | - | 1 | 3 | 0 | 1 | 3 |
| | SDG 4.1.1b | - | 2 | 2 | - | 2 | 2 | - | 2 | - |
| | SDG 4.1.1c | - | 4 | - | - | 3 | - | - | 2 | - |
| | SDG 4.1.1a & b | - | - | - | - | - | - | - | - | 1 |
| | SDG 4.1.1b & c | 4 | 2 | - | 4 | 3 | - | 4 | 4 | 2 |
| | SDG 4.1.1a & c | - | - | - | 1 | - | - | - | - | 1 |
| | SDG 4.1.1a, b & c | - | - | 1 | - | - | 1 | - | - | 1 |
| ECA | Don't report on SDG 4.1.1a, b & c | 2 | 2 | 8 | 2 | 2 | 8 | 2 | 2 | 8 |
| | SDG 4.1.1a | - | 1 | 1 | - | - | 1 | - | - | 1 |
| | SDG 4.1.1b | - | 1 | 1 | - | 1 | 1 | - | 1 | - |
| | SDG 4.1.1c | - | 5 | 4 | 3 | 5 | - | 3 | 5 | - |
| | SDG 4.1.1a & b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1b & c | 19 | 8 | 2 | 20 | 8 | 2 | 20 | 8 | 3 |
| | SDG 4.1.1a & c | 3 | - | - | 3 | - | - | 3 | - | - |
| | SDG 4.1.1a, b & c | 1 | - | - | 1 | 1 | - | 1 | 1 | - |
| LAC | Don't report on SDG 4.1.1a, b & c | 2 | 11 | 9 | 2 | 10 | 8 | 2 | 10 | 8 |
| | SDG 4.1.1a | - | 1 | - | - | 2 | - | - | 2 | - |
| | SDG 4.1.1b | - | - | - | - | - | 1 | - | - | - |
| | SDG 4.1.1c | - | 1 | - | - | 1 | - | - | 1 | - |
| | SDG 4.1.1a & b | 1 | 14 | 2 | 1 | 14 | 2 | 1 | 14 | 2 |
| | SDG 4.1.1b & c | - | - | - | - | - | - | - | - | 1 |
| | SDG 4.1.1a & c | - | 1 | - | - | 1 | - | - | 1 | - |
| | SDG 4.1.1a, b & c | - | 1 | - | - | 1 | - | - | 1 | - |
| MNA | Don't report on SDG 4.1.1a, b & c | - | 4 | 3 | - | 4 | 3 | - | 4 | 2 |
| | SDG 4.1.1a | - | - | 1 | - | - | 1 | - | - | - |
| | SDG 4.1.1b | - | 1 | - | - | 1 | - | - | - | - |
| | SDG 4.1.1c | 1 | 1 | - | - | 1 | - | 1 | 1 | - |
| | SDG 4.1.1a & b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1b & c | 5 | 3 | 2 | 6 | 3 | 2 | 6 | 4 | 2 |
| | SDG 4.1.1a & c | - | - | - | - | - | - | - | - | 1 |
| | SDG 4.1.1a, b & c | - | - | - | - | - | - | - | - | - |
| SAR | Don't report on SDG 4.1.1a, b & c | - | 4 | 3 | - | 4 | 3 | - | 4 | 2 |
| | SDG 4.1.1a | - | - | 1 | - | - | 1 | - | - | - |
| | SDG 4.1.1b | - | 1 | - | - | 1 | - | - | - | - |
| | SDG 4.1.1c | 1 | 1 | - | - | 1 | - | 1 | 1 | - |
| | SDG 4.1.1a & b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1b & c | 5 | 3 | 2 | 6 | 3 | 2 | 6 | 4 | 2 |
| | SDG 4.1.1a & c | - | - | - | - | - | - | - | - | 1 |
| | SDG 4.1.1a, b & c | - | - | - | - | - | - | - | 2 | - |
| SSA | Don't report on SDG 4.1.1a, b & c | - | 19 | 10 | - | 15 | 9 | - | 15 | 7 |
| | SDG 4.1.1a | - | 3 | 1 | - | 2 | 6 | - | 2 | 6 |
| | SDG 4.1.1b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1c | - | - | - | - | - | - | - | - | 1 |
| | SDG 4.1.1a & b | - | 11 | 4 | - | 11 | 4 | - | 10 | 4 |

| | Col1 | Col2 | Col3 | Col4 | Col5 | Col6 | Col7 | Col8 | Col9 |
|---|---|---|---|---|---|---|---|---|---|
| SDG 4.1.1b & c | - | - | - | - | 1 | - | - | 1 | - |
| SDG 4.1.1a & c | - | - | - | - | - | - | - | - | - |
| SDG 4.1.1a, b **& c** | - | - | - | - | - | - | 1 | 1 | - |

Note: SDG 4.1.1a reporting for high-income countries tends to rely on PIRLS grade 4 whereas other countries use assessments done at grades 2/3.

## Table A4. SDG reporting coverage by CMU

| | | Reporting on SDG 4.1.1 over last 3 years (2018 or later) | | | Reporting on SDG 4.1.1 over last 5 years (2016 or later) | | | Reporting on SDG 4.1.1 over last 7 years (2014 or later) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Lower secondary enrolled >= 90% | Lower secondary enrolled < 90% | No enrollment data | Lower secondary enrolled >= 90% | Lower secondary enrolled < 90% | No enrollment data | Lower secondary enrolled >= 90% | Lower secondary enrolled < 90% | No enrollment data |
| **AFR CW 1** | Don't report on SDG 4.1.1a, b **& c** | - | 3 | 1 | - | 3 | 1 | - | 3 | 1 |
| | SDG 4.1.1a | - | - | 1 | - | - | 1 | - | - | 1 |
| | SDG 4.1.1b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a & b | - | 8 | - | - | 8 | - | - | 7 | - |
| | SDG 4.1.1b & c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a & c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a, b **& c** | - | - | - | - | - | - | - | 1 | - |
| **AFR CW 2** | Don't report on SDG 4.1.1a, b **& c** | - | 4 | 2 | - | 2 | 1 | - | 2 | 1 |
| | SDG 4.1.1a | - | - | - | - | 2 | 1 | - | 2 | 1 |
| | SDG 4.1.1b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a & b | - | 1 | 2 | - | 1 | 2 | - | 1 | 2 |
| | SDG 4.1.1b & c | - | - | - | | | | | | |
| | SDG 4.1.1a & c | - | - | - | | | | | | |
| | SDG 4.1.1a, b **& c** | - | - | - | - | - | - | - | - | - |
| **AFR ES 1** | Don't report on SDG 4.1.1a, b **& c** | - | 9 | 2 | - | 7 | 2 | - | 7 | 1 |
| | SDG 4.1.1a | - | 1 | - | - | 2 | - | - | 2 | - |
| | SDG 4.1.1b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1c | - | - | - | - | - | - | - | - | 1 |
| | SDG 4.1.1a & b | - | 1 | - | - | 1 | - | - | 1 | - |
| | SDG 4.1.1b & c | - | - | - | - | 1 | - | - | 1 | - |
| | SDG 4.1.1a & c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a, b **& c** | - | - | - | - | - | - | - | - | - |
| **AFR ES 2** | Don't report on SDG 4.1.1a, b **& c** | - | 3 | 5 | - | 3 | 5 | - | 3 | 4 |
| | SDG 4.1.1a | - | 2 | - | - | 2 | - | - | 2 | - |
| | SDG 4.1.1b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1c | - | 1 | 2 | - | 1 | 2 | - | - | - |
| | SDG 4.1.1a & b | - | - | - | - | - | - | - | 1 | 2 |
| | SDG 4.1.1b & c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a & c | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1a, b **& c** | - | - | - | - | - | - | - | - | 1 |
| **EAP** | Don't report on SDG 4.1.1a, b **& c** | - | 7 | 12 | - | 7 | 10 | - | 7 | 9 |
| | SDG 4.1.1a | - | 1 | 2 | - | 1 | 3 | 0 | 1 | 3 |
| | SDG 4.1.1b | - | 2 | 2 | - | 2 | 2 | - | 2 | - |
| | SDG 4.1.1c | - | 4 | - | - | 3 | - | - | 2 | - |
| | SDG 4.1.1a & b | - | - | - | - | - | - | - | - | 1 |

| Region | Category | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SDG 4.1.1b & c | 4 | 2 | - | 4 | 3 | - | 4 | 4 | 2 |
| | SDG 4.1.1a & c | - | - | - | 1 | - | - | - | - | 1 |
| | SDG 4.1.1a, b & c | - | - | 1 | - | - | 1 | - | - | 1 |
| ECA | Don't report on SDG 4.1.1a, b & c | 2 | 2 | 8 | 2 | 2 | 8 | 2 | 2 | 8 |
| | SDG 4.1.1a | - | 1 | 1 | - | - | 1 | - | - | 1 |
| | SDG 4.1.1b | - | 1 | 1 | - | 1 | 1 | - | 1 | - |
| | SDG 4.1.1c | - | 5 | 4 | 3 | 5 | - | 3 | 5 | - |
| | SDG 4.1.1a & b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1b & c | 19 | 8 | 2 | 20 | 8 | 2 | 20 | 8 | 3 |
| | SDG 4.1.1a & c | 3 | - | - | 3 | - | - | 3 | - | - |
| | SDG 4.1.1a, b & c | 1 | - | - | 1 | 1 | - | 1 | 1 | - |
| LCR | Don't report on SDG 4.1.1a, b & c | 2 | 11 | 9 | 2 | 10 | 8 | 2 | 10 | 8 |
| | SDG 4.1.1a | - | 1 | - | - | 2 | - | - | 2 | - |
| | SDG 4.1.1b | - | - | - | - | - | 1 | - | - | - |
| | SDG 4.1.1c | - | 1 | - | - | 1 | - | - | 1 | - |
| | SDG 4.1.1a & b | 1 | 14 | 2 | 1 | 14 | 2 | 1 | 14 | 2 |
| | SDG 4.1.1b & c | - | - | - | - | - | - | - | - | 1 |
| | SDG 4.1.1a & c | - | 1 | - | - | 1 | - | - | 1 | - |
| | SDG 4.1.1a, b & c | - | 1 | - | - | 1 | - | - | 1 | - |
| MNA | Don't report on SDG 4.1.1a, b & c | - | 4 | 3 | - | 4 | 3 | - | 4 | 2 |
| | SDG 4.1.1a | - | - | 1 | - | - | 1 | - | - | - |
| | SDG 4.1.1b | - | 1 | - | - | 1 | - | - | - | - |
| | SDG 4.1.1c | 1 | 1 | - | - | 1 | - | 1 | 1 | - |
| | SDG 4.1.1a & b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1b & c | 5 | 3 | 2 | 6 | 3 | 2 | 6 | 4 | 2 |
| | SDG 4.1.1a & c | - | - | - | - | - | - | - | - | 1 |
| | SDG 4.1.1a, b & c | - | - | - | - | - | - | - | - | - |
| SAR | Don't report on SDG 4.1.1a, b & c | - | 4 | 3 | - | 4 | 3 | - | 4 | 2 |
| | SDG 4.1.1a | - | - | 1 | - | - | 1 | - | - | - |
| | SDG 4.1.1b | - | 1 | - | - | 1 | - | - | - | - |
| | SDG 4.1.1c | 1 | 1 | - | - | 1 | - | 1 | 1 | - |
| | SDG 4.1.1a & b | - | - | - | - | - | - | - | - | - |
| | SDG 4.1.1b & c | 5 | 3 | 2 | 6 | 3 | 2 | 6 | 4 | 2 |
| | SDG 4.1.1a & c | - | - | - | - | - | - | - | - | 1 |
| | SDG 4.1.1a, b & c | - | - | - | - | - | - | - | 2 | - |

Note: SDG 4.1.1a reporting for high-income countries tends to rely on PIRLS grade 4 whereas other countries use assessments done at grades 2/3.

**Table A5. Lending Group and GPE Classifications**

| Lending Group | Not GPE | GPE | Totals |
|---|---|---|---|
| HIC | 69 | 0 | **69** |
| IBRD | 54 | 16 | **70** |
| IDA/Blend | 1 | 73 | **74** |
| Not Classified | 3 | 1 | **4** |
| **Total** | **127** | **90** | **217** |

Table A6. Assessment costs under different simulation scenarios

| | Scenario 1: Low | Scenario 2: Intermediate | Scenario 3: High | Scenarios 1-3: Custom Group 3 Proportions | Scenario 4: Custom Country Inputs | Countries |
|---|---|---|---|---|---|---|
| *Total* | $ 196,552,140 | $ 227,037,523 | $ 259,732,140 | $ 227,037,523 | $ 66,194,238 | 132 |
| *Region* | | | | | | |
| EAP | $ 42,453,598 | $ 48,062,582 | $ 55,413,599 | $ 48,062,582 | $ 15,102,425 | 30 |
| ECA | $ 39,331,983 | $ 44,755,120 | $ 49,051,983 | $ 44,755,120 | $ 19,250,538 | 25 |
| LAC | $ 28,626,299 | $ 35,596,265 | $ 44,826,299 | $ 35,596,265 | $ 3,884,713 | 24 |
| MNA | $ 16,285,477 | $ 18,997,045 | $ 21,145,477 | $ 18,997,045 | $ 7,006,425 | 11 |
| NAC | $ 1,227,856 | $ 1,679,784 | $ 2,037,856 | $ 1,679,784 | $ - | 1 |
| SAS | $ 20,038,369 | $ 20,490,297 | $ 20,848,369 | $ 20,490,297 | $ 13,156,525 | 8 |
| SSA | $ 48,588,558 | $ 57,456,430 | $ 66,408,558 | $ 57,456,430 | $ 7,793,613 | 33 |
| **Total** | **$ 196,552,140** | **$ 227,037,523** | **$ 259,732,140** | **$ 227,037,523** | **$ 66,194,238** | **132** |
| *Lending Group* | | | | | | |
| IDA/Blend | $ 89,131,551 | $ 103,236,806 | $ 118,291,551 | $ 103,236,806 | $ 20,767,572 | 58 |
| IBRD | $ 58,164,707 | $ 63,985,275 | $ 70,314,707 | $ 63,985,275 | $ 27,920,466 | 37 |
| HIC | $ 43,315,046 | $ 52,620,190 | $ 62,755,046 | $ 52,620,190 | $ 17,506,200 | 34 |
| GPE only | $ 1,227,856 | $ 1,679,784 | $ 2,037,856 | $ 1,679,784 | $ - | 1 |
| Not classified | $ 3,485,123 | $ 3,835,684 | $ 4,295,123 | $ 3,835,684 | $ - | 2 |
| **Total** | **$ 195,324,284** | **$ 225,357,739** | **$ 257,694,284** | **$ 225,357,739** | **$ 66,194,238** | **132** |
| *Income* | | | | | | |
| LIC | $ 33,021,437 | $ 39,159,655 | $ 45,171,437 | $ 39,159,655 | $ 4,763,000 | 21 |
| LMC | $ 66,284,978 | $ 71,461,518 | $ 76,814,978 | $ 71,461,518 | $ 28,354,413 | 38 |
| UMC | $ 41,285,089 | $ 48,494,352 | $ 56,675,089 | $ 48,494,352 | $ 11,092,813 | 30 |
| HIC | $ 55,960,637 | $ 67,921,998 | $ 81,070,637 | $ 67,921,998 | $ 21,984,013 | 43 |
| **Total** | **$ 196,552,140** | **$ 227,037,523** | **$ 259,732,140** | **$ 227,037,523** | **$ 66,194,238** | **132** |
| *CMU (IDA/Blend/IBRD)* | | | | | | |
| AFRCW1 | $ 6,980,681 | $ 8,593,021 | $ 10,220,681 | $ 8,593,021 | $ - | 5 |
| AFRCW2 | $ 10,740,510 | $ 11,949,766 | $ 13,170,510 | $ 11,949,766 | $ 1,350,000 | 6 |
| AFRES1 | $ 15,680,151 | $ 18,904,832 | $ 22,160,151 | $ 18,904,832 | $ 3,900,600 | 12 |
| AFRES2 | $ 15,187,216 | $ 18,008,812 | $ 20,857,216 | $ 18,008,812 | $ 2,543,013 | 10 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *EAP* | $ 29,948,731 | $ 33,454,346 | $ 38,048,731 | $ 33,454,346 | $ 11,210,125 | 21 |
| *ECA* | $ 17,391,988 | $ 18,747,773 | $ 19,821,988 | $ 18,747,773 | $ 7,585,838 | 10 |
| *LCR* | $ 18,416,332 | $ 21,901,315 | $ 26,516,332 | $ 21,901,315 | $ 3,884,713 | 14 |
| *MNA* | $ 12,912,281 | $ 15,171,921 | $ 16,962,281 | $ 15,171,921 | $ 5,057,225 | 9 |
| *SAR* | $ 20,038,369 | $ 20,490,297 | $ 20,848,369 | $ 20,490,297 | $ 13,156,525 | 8 |
| ***Total*** | **$ 147,296,259** | **$ 167,222,081** | **$ 188,606,259** | **$ 167,222,081** | **$ 48,688,038** | ***95*** |

*Note. The numbers above are for countries in Group 2a, 2b, and 3. Scenario 1 denotes 50% policy or student linking only, 16.6% NLA + policy or student linking, regional, and international each; Scenario 2 denotes 25% policy or student linking only, NLA + policy or student linking, regional, and international each; Scenario 3 denotes 16.6% policy or student linking only, 50% NLA + policy or student linking, and 16.6% regional, and international each. Scenarios 1-3 are customizable in the accompanying Excel tool and currently depict results for the "intermediate" scenario. Scenario 4 is custom based on user's country-level inputs in sheet "input_table" and currently display "default" baseline costs based on existing assessments.*

## Figure A1. Theory of Change for the LDC at a National Level

**Figure A2. Theory of Change for the LDC at a Global Level**

**Global**

| Inputs | Activities | Outputs | Outcomes I | Outcomes II | Impact |
|--------|-----------|---------|-----------|------------|--------|
| *What is needed?* | *What is done?* | *What is produced immediately?* | *What are the short- to medium-term results?* | *What are the medium-term results?* | *What are the long-term results?* |

| Inputs | Activities | Outputs | Outcomes I | Outcomes II | Impact |
|--------|-----------|---------|-----------|------------|--------|
| Adequate staff<br><br>Adequate funds<br><br>Adequate time<br><br>Adequate managerial support from partner institutions<br><br>Adequate support from development partners and donors | **Convening Activities**<br>• Presentations<br>• Workshops<br>• Meetings<br>• Technical secretariat<br>• Advisory board<br><br>**Knowledge Activities**<br>• Creation of new technical products<br>• Dissemination of knowledge (TA)<br><br>**Financing Activities**<br>• Disbursement of country grants and loans<br>• Domestic resource mobilization | Agreement on priorities and principles of the LDC<br><br>Shared country-level mapping of LDC-related partner activities<br><br>Knowledge exchanged through LDC networks<br><br>Global public goods, such as global item banks, knowledge products, and technical resources | Improved alignment and coordination on key priorities among LDC partners<br><br>Less duplication and fragmentation of resources<br><br>Uptake and dissemination of global public goods and technical resources | Shared protocols among the LDC partners<br><br>Aligned support to countries on data collection and use through stronger collaborations<br><br>Improved oversight and transparency of global efforts to measure student learning outcomes<br><br>Stronger support to efforts complementing regional and international assessment initiatives | Availability of regular, timely, comparable, policy-relevant, country-owned learning data<br><br>Stronger global efforts to improve student learning outcomes |

*Feedback loops*

**Risk factors:** Poor institutional coordination, competing initiatives and brands, lack of in-country demand, lack of funding

**Figure A3. Country group classifications**

| | 3 years | 5 years | 7 years | N years |
|---|---|---|---|---|
| 2 or more SDG measurement points reported | (green) | (orange) | (orange) | (red) |
| 1 SDG measurement point reported | (yellow) | (yellow) | (yellow) | (red) |
| 0 SDG measurement points reported | (red) | (red) | (red) | (red) |

**Section A1. Simulation Notes**

**Calculation notes for simulation scenarios 1-3**

- Identify the most recent assessments used to report on SDG 4.1.1. a, b, and c respectively over the last 7 years (2014 onwards).

- Calculate the cost of assessment to report on SDG 4.1.1. a, b, and c each respectively. We use the cost of assessment that is standardized to reflect the cost for covering one grade and one subject. We use the cost of NLA without any policy or student linking.

- Sum across the cost of assessments for the 3 SDG measurement points to get baseline costs for the most recent assessments over the last 7 years.

- We don't add any additional costs for Group 1 and 2a since they have already reported on 2 or more SDG measurement points.

- Add additional costs for Group 2b, so that they reflect two grades and two subjects. This entails:

    o For Group 2b, additional cost of doing 1 more subject and grade – baseline costs only reflect 1 subject and 1 grade. We add the adjusted unit cost of assessment for 1 grade and 1 subject.

- Add additional costs for Group 3 for covering 2 grades and 2 subjects under each of the following "extreme" scenarios: all Group 3 countries only do policy or student linking, or NLA and policy or student linking, or regional assessment, or international assessment.

- Add the baseline cost and all the additional costs for Groups 2b and 3 to calculate the total costs to achieve the goal of 2 x 2 x 1 under each of the "extreme" scenarios.

- We use the total cost for Group 3 under each of these "extreme" scenarios to calculate the cost if only a certain proportion of Group 3 countries do a specific type of assessment. So under the "intermediate" scenario, the calculation for total cost is as follows: add 25% of the total cost for Group 3 under "extreme" scenario where all Group 3 countries do policy or student linking, 25% of the total cost for Group 3 under "extreme" scenario where all Group 3 countries do NLA and policy or student linking, 25% of the total cost for Group 3 under "extreme" scenario where all Group 3 countries do regional assessments, and 25% of the total cost for Group 3 where all Group 3 countries do international assessments, as well as the "static" costs for the remaining groups 1, 2a, and 2b to attain 2 x 2 x 1 goals. This produces the simulated costs under scenarios 1-3 described in earlier sections.

**Calculation notes for simulation scenario 4**

- Identify the two or less most recent assessments used by the country to report on SDG 4.1.1a, b, or c in the last 7 years.

- The "Default" selection gives an estimate of the existing baseline cost of assessments.

    o Where available, the "Default" selection chooses the unadjusted costs for assessments 1 and 2. In instances where both assessments are NLAs, we assume neither include policy or student linking costs. We use the unadjusted cost of NLA only for the first assessment, and the adjusted cost of 1 additional grade *or* 1 additional subject for the second assessment. Note: NLAs may be under counted in cases where they have been reported

for more than 2 SDG 4.1.1 measurement points, or NLA costs may be underestimated where countries have used them for policy or student linking.

- o For countries where both assessments are TIMSS, or that have reported on two SDGs using TIMSS, we use the unadjusted cost of two subjects and two grades for one assessment and don't reflect its cost for assessment 2. For countries that have reported on only one SDG with TIMSS, we also count it as 1 assessment and use the adjusted cost of one subject and one grade. In either case, assessment 2 costs can reflect a different assessment (such as PISA) if a country has reported more than 1 SDGs using TIMSS + another assessment.

- o For countries where the assessment was MICS, we assume the full cost of administering a household survey and a foundational learning module.

- If the selection is "Custom," and:

  - o If the first assessment chosen by the user is already for 2 grades and 2 subjects (for example, ERCE, PILNA, PASEC, or TIMSS), the user only has the option to add PISA if they would still like to choose a second assessment. We use the unadjusted actual cost for both assessments.

  - o Similarly, users will only have the option to choose a second assessment that is 2 grades and 2 subjects if the first assessment selection is PISA.

  - o If the first assessment chosen by the user is for 1 grade and 1 subject (for example, PIRLS), and the second assessment is the same (PIRLS), we apply an adjusted value for the costs (for 1 grade *and* 1 subject) of the second assessment but an unadjusted value for the cost of the first assessment. If the second assessment is different from the first assessment, we use the unadjusted value for both assessments.

  - o If the first assessment chosen by the user if for 1 grade and 2 subjects (for example, PISA, PISA4D, SACMEQ, SEA-PLM, MICS, PAL Network assessments, national assessments, MILO, or EGRA), and the second assessment is the same, we apply an adjusted value (for 1 additional grade *or* 1 additional subject) for the cost of the second assessment but an unadjusted value for the cost of the first assessment. If the second assessment is different from the first assessment, we use the unadjusted value for both assessments.

  - o If the first assessment chosen by the user is MICS, users can choose whether the country had a prior MICS survey, a prior household survey, or no surveys. If the second assessment is also MICS, the discounted cost for the same option will be applied.

  - o If the first assessment is a new national assessment, users can choose to administer tests with or without policy/student linking costs. If the first assessment is an existing national assessment, the cost of only policy or student linking is applied. If the first assessment is a national assessment, the second assessment cannot be a new national assessment, but could be policy/student linking only, or another assessment. If a user chooses a national assessment for the second assessment but not the first, it will be treated as a new assessment and will also have the option to include policy/student linking.

**End Notes**

[1] World Bank 2021.

[2] Azevedo et al (2021a).

[3] World Bank 2019.

[4] Azevedo et al (2021b) and Azevedo (2020).

[5] Both approaches have been developed in close collaboration with governments from different regions, such as India, Zambia, Ghana, Morocco, Bangladesh, Lesotho, Nigeria, Cambodia, Guatemala, Colombia, Burkina Faso, Senegal, Côte d'Ivoire, with regional partners such as PASEC and ERCE, and with global partners such as UIS, USAID, the World Bank, IEA, and ACER.

[6] Clarke and Bazaldua, 2021

[7] Hoyos, Garcia-Moreno, and Patrinos 2017.

[8] Hoyos, Ganimian, and Holland 2017

[9] Bruns, Akmal, and Birdsall 2019

[10]
https://www.worldbank.org/en/publication/wdr2018

[11] http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/Background/EMIS/Framework_SABER-EMIS.pdf

[12] Clarke 2015; Results for Development 2015

[13] https://www.unicef.org/media/63056/file/EdStrategy-2019-2030-CountrySolution-WCAR.pdf

[14] Available methologies to be implemented were developed and applied by IIEP UNESCO Dakar

[15] https://www.unicef-irc.org/research/time-to-teach/

[16] One potential scenario is different business models for global and regional global goods and country funding where country funding goes to countries while public goods are funded from a common fund.

[17] This note accompanies the Excel tool containing the data and costing simulations. The Excel tool can be accessed here.

[18] The cost simulations below are for one round of assessment only.

[19] SDG Indicator 4.1.1: Proportion of children and young people (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex.

[20] See the "Data and Methodology" section for details on methodology used to categorize countries into different groups, and the number of countries in each group.

[21] The total and IBRD baseline costs represents the estimate from simulation 4 (custom) and the lower end represents the estimate from simulations 1-3 (low, intermediate, high). For IDA/Blend, the upper end represents the estimate from simulations 1-3 and the lower end represents the estimate from simulation 4. The two sets of simulations rely on different computations to calculate baseline costs, with simulation 4 largely relying on unadjusted total costs for the two recent assessments identified, and simulations 1-3 relying on adjusted costs for 1 grade and 1 subject for the assessments identified at each of the three SDG 4.1.1 measurement points. See more details on the cost calculation methodology in the Appendix.

[22] Note that *all* countries will be able to benefit from the global public goods produced by the LDC.

[23] Assessments such as SEA-PLM, PIRLS, PISA, and LaNA only cover one grade. Most other assessments cover two grades.

[24] Group 1 countries have reported on 2 or more SDG measurement points through assessments that took place in the last 3 years. Group 2a countries have reported on 2 or more SDG measurement points through assessments that took place more than 3 years ago but less than 7 years ago. Group 2b countries have only reported on 1 SDG measurement point through assessments that took place over the last 7 years. Group 3 countries have reported on either no SDG measurement points or have reported on one or more SDG measurement point through an assessment that is older than 7 years.

[25] Last three years, or between 2018 and 2020. Tables A1 and A2 in the annex have the same results using a 5 and 7 year intervals.

[26] GPE countries as listed here on July 21, 2021: https://www.globalpartnership.org/where-we-work/partner-countries

[27] Only one IDA/Blend country is not a GPE country: Kosovo.

[28] We focus our analysis on the 217 World Bank countries, and exclude 23 territories only available in UIS SDG database from this analysis: Saint Pierre and Miquelon, Anguilla, Aland Islands, Saint Barthélemy, Cook Islands, Western Sahara, Falkland Island, Guernsey, Guadeloupe, French Guiana, Jersey, Montserrat, Martinique, Mayotte, Norfolk Island, Niue, Pitcairn, Reunion, Saint Helena, Svalbard and Jan Mayan, Tokelau, Holy See, and Wallis and Futura Islands.

[29] There were 62 out of 217 countries with unavailable enrollment data after 2013: Algeria, American Samoa, Andorra, Angola, Armenia, Aruba, Bermuda, Bosnia and Herzegovina, Botswana, Cambodia, Cayman Islands, Channel Islands, China, Congo, Dem. Rep., Congo, Rep., Curacao, Equatorial Guinea, Faroe Islands, Fiji, French Polynesia, Gabon, Gambia, The, Greenland, Guam, Guinea-Bissau, Guyana, Haiti, Iraq, Isle of Man, Japan, Kenya, Kiribati, Korea, Dem People's Rep, Kosovo, Lebanon, Libya, Maldives, Micronesia, Fed Sts, Monaco, Mongolia, Namibia, New Caledonia, Nicaragua, Nigeria, North Macedonia, Northern Mariana Islands, Palau, Paraguay, Qatar, Solomon Islands, Somalia, Martin (French part), Sudan, Tajikistan, Trinidad and Tobago, Tunisia, Turkmenistan, Turks and Caicos Islands, Uganda, Vietnam, Virgin Islands (US), Zambia.

[30] The average reporting year for enrollment is 2017. The average reporting year is 2017.0 for 61 available high-income countries, 2017.4 for 38 available upper-middle income countries, 2016.9 for 36 available lower-middle income countries, and 2016.6 for 20 available low-income countries.