CHAPTER 29

Government Analytics Using Measures of Service Delivery

Kathryn Andrews, Galileu Kim, Halsey Rogers, Jigyasa Sharma, and Sergio Venegas Marin

SUMMARY

Public services, such as primary health care and education, have important consequences for social welfare and economic development. However, the quality of service delivery across the world is uneven. To improve it, practitioners require evidence to understand what is driving outcomes in education and health, such as student learning and the prevalence of chronic diseases. *Measures of service delivery* (MSDs) provide objective measurements of the quality of public service delivery. These indicators offer a granular view of the service delivery system, providing actionable insights into different parts of the delivery chain: from the physical infrastructure to the knowledge of frontline providers. This chapter provides an outline for how to conceptualize, measure, and disseminate MSDs, leveraging the institutional expertise of teams of practitioners at the World Bank. It offers actionable steps and advice that aim to connect practitioners to wider global efforts to improve the quality of public service delivery.

ANALYTICS IN PRACTICE

- *Measures of service delivery* (MSDs) are objective measures of different parts of the public service delivery system. These indicators provide a granular view of the entire process of service delivery. MSDs measure the quality of the delivery of public services, such as primary health care and education. In addition to measuring welfare outcomes, measures should focus on different parts of the service delivery system, such as physical capital (for example, hospitals and schools) and human capital (for example, the knowledge of practitioners). Management practices play an important role in translating physical infrastructure and human capital into patient and student outcomes.

This chapter's authors are World Bank staff. Kathryn Andrews is a health economist in the Health, Nutrition, and Population Global Practice. Galileu Kim is a research analyst in the Development Impact Evaluation (DIME) Department. Halsey Rogers is a lead economist in the Education Global Practice. Jigyasa Sharma is a health economist in the Health, Nutrition, and Population Global Practice. Sergio Venegas Marin is an education economist in the Education Global Practice.

- Designing MSDs allows practitioners to specify the dimensions of service quality and construct measures to identify how well services perform in each one of them. Developing MSDs for primary health care and education requires considering and defining what dimensions will be used to measure the quality of public services. For example, a personnel dimension may measure absence rates for teachers and doctors.[1] Another dimension may be the availability of learning materials in schools and medical supplies in health facilities. There is a variety of conceptual frameworks and indicators to draw from. For health MSDs, practitioners may build on the existing frameworks described by *The Lancet Global Health* Commission (Hanson et al. 2022) or the World Health Organization's "building blocks" framework (WHO 2010).

- The implementation of MSDs should follow a sequential structure, from defining a conceptual framework around measurements of the quality of public services to disseminating findings to government stakeholders and citizens. Generally, the first step in the implementation process is defining a conceptual framework and securing institutional support. The next step is identifying what will be measured: the indicators of interest; the questions to be asked; and to whom, where, and with what frequency they will be asked. After defining these indicators, practitioners should develop a rollout strategy for the actual data collection, which could include procuring a survey firm or developing a specific management information system for health care or education. After the data are collected, they should be validated, processed, and transformed into indicators. The final step is crucial: the resulting MSDs should be clearly articulated to stakeholders and disseminated widely, both within the government and to citizens.

- While practitioners may develop MSDs independently, the development of objective measurements to improve public services is part of a global agenda. Connecting to this broader movement allows practitioners to learn from other governments' experiences. Engagement with global partners can also accelerate the design and implementation of MSDs. This global engagement can raise awareness of the relative standing of countries through benchmarking exercises, as well as facilitate knowledge exchange.

- MSDs should be subject to constant revision, as the understanding of quality in service delivery evolves. Indicators should also respond to new and unexpected demands. MSDs should evolve according to the changing policy objectives of stakeholders and citizens. Adaptations in measurement methodologies reflect an ongoing dialogue between policy makers, citizens, and the practitioners responsible for producing these indicators. As the COVID-19 (coronavirus) pandemic has highlighted, moments of crisis may generate demand for additional indicators, such as the availability of vaccines and the impact of school shutdowns on student learning.

INTRODUCTION

Governments are responsible for the delivery of public services in primary health care and education, the foundations of public health and student learning.[2] It is well established in the development community that these services have immediate and important consequences for citizens who depend on them. Children rely on education services to learn how to read and write (World Bank 2018), and, as the COVID-19 pandemic has highlighted, access to health services can often determine whether patients recover from severe infections (Gatti et al. 2021). However, these same reports highlight the uneven quality of these services (Andrews et al. 2021). While some citizens may receive high-quality services, with positive impacts on welfare outcomes such as learning and health, others do not. What can practitioners do to improve the quality of these services? How can one measure changes in the quality of service? And what policy levers are available to change them?

We first note that public services are the outputs of a complex *service delivery system*. This system includes a range of inputs and processes. First, policies define how the service delivery system is structured and accessed: policies prescribe who is eligible to receive these services and establish processes to select service providers, such as teachers and nurses. These de jure policies lay the institutional groundwork for service delivery, but de facto inputs are equally important. These inputs include the facilities necessary for the provision of these services, such as schools and clinics, and the materials necessary for daily operations. Human capital is also crucial: these are the practitioners responsible for teaching students, for diagnosis and treatment, and for using their knowledge and skills to provide these services. Finally, a range of processes and management practices—including referral systems, feedback mechanisms, and counseling—translate these physical and human resources into welfare outcomes.[3]

Measures of service delivery (MSDs) measure the quality of these different dimensions of service delivery. MSDs account for and measure multiple factors in the service delivery system, providing policy makers with a holistic and granular view of how public services operate. Measurement of these different factors of production allows practitioners to map out conceptually how each part of the production chain is faring and where improvements can be made (Amin, Das, and Goldstein 2007). MSDs not only allow practitioners to measure each part of the chain; they also uncover causal relationships. As noted by Amin, Das, and Goldstein (2007), one of the key contributions of MSDs is allowing practitioners and researchers to measure the impact of a policy intervention in a rigorous way. Given these potential benefits, how can practitioners develop these indicators?

In this chapter, we focus on two examples of MSDs: the Service Delivery Indicators (SDI) Health Survey and the Global Education Policy Dashboard (GEPD). Drawing on both teams' expertise, we present an overview to practitioners on how to develop MSDs for primary health care and education, focusing on the facilities (schools and health facilities) in which these services are provided. Given this scope, we acknowledge there are secondary and even tertiary levels and other public services that stand to benefit from better measurement (for example, social protection and transportation) and that the private sector often provides these crucial services as well.[4] Nonetheless, we hope this chapter serves as an applied example for government practitioners on how to develop, implement, and use MSDs to improve the quality of public services at a foundational level. We outline conceptual frameworks and indicators, as well as how to generate them from survey data. Additionally, we highlight the benefits of connecting to a global agenda to develop and improve these indicators of service delivery.

This chapter is structured as follows. First, we provide a conceptual framework to measure the quality of service delivery. Section three outlines practical steps for implementation, adapted from the experience of our practitioners. Section four outlines the broader global agenda for developing MSDs. Finally, we conclude.

AN OVERVIEW OF MEASURES OF SERVICE DELIVERY

Multiple global initiatives promote the use of MSDs. These include the World Bank's Learning Poverty indicator (World Bank 2021) and the Primary Health Care Performance Initiative (PHCPI).[5] Table 29.1 provides an abridged list of key global initiatives in generating MSDs, highlighting other related measurement initiatives. For example, the World Health Organization (WHO) Service Availability and Readiness Assessment (SARA) and the United States Agency for International Development (USAID) Service Provision Assessment (SPA) are similar initiatives to the SDI Health Survey. The revamped SDI Health Survey draws on best practices from both the SPA and SARA and goes a step further in the comprehensiveness of its domains of measurement and its patient-centered focus.

TABLE 29.1 **Survey of Global Initiatives in Education and Health Care Delivery Indicators**

Public service	Initiative	Description
Education	Global Education Policy Dashboard (GEPD)	To help countries put an end to learning poverty, the World Bank's Education Global Practice has developed and is supporting countries in the deployment of the GEPD. This new tool offers a strong basis for identifying priorities for investment and policy reforms that are suited to each country's context. It does so by highlighting gaps between what the evidence suggests is effective in promoting learning and what is happening in practice in each system and allowing governments to track progress as they take action to close those gaps.
	Service Delivery Indicators (SDI) Education	The SDI Education initiative collects data on service delivery in school facilities. It helps countries identify areas of progress and areas for improvement with potential lessons for progress within and between countries. Collected in close collaboration with the countries requesting a diagnostic, the data are used to assess the quality and performance of education. Since the initiative's creation, the surveys used have evolved, and existing data sets have been harmonized to allow for country comparisons over time.
	Systems Approach for Better Education Results Service Delivery (SABER SD)	The SABER SD tool was developed in 2016, in the Global Engagement and Knowledge Unit of the Education Global Practice at the World Bank, as an initiative to uncover bottlenecks that inhibit student learning in low- and middle-income countries and to better understand the quality of education service delivery in countries, as well as gaps in policy implementation. This school survey is aligned with the latest education research on what matters for student learning and how best to measure it. Its main purposes are to provide a mechanism to assess different determinants of learning through a diagnostic tool and to uncover the extent to which policies translate into practice.
	Teach Early Childhood Education (ECE)	Teach ECE is a free classroom observation tool that provides a window into one of the less explored and more important aspects of a child's education: what goes on in the classroom. The tool is intended to be used with children ages three to six and was designed to help countries, in particular low- and middle-income countries, monitor and improve teaching quality following the Teach Primary framework.
	Learning Poverty indicator	This indicator brings together schooling and learning indicators: it begins with the share of children who haven't achieved minimum reading proficiency (as measured in schools) and is adjusted by the proportion of children who are out of school (and are assumed not to be able to read proficiently).
	COVID-19: Monitoring the Impacts on Learning Outcomes (MILO)	The MILO project aims to measure learning outcomes in six countries in Africa in order to analyze the long-term impact of COVID-19 on learning and to evaluate the effectiveness of distance-learning mechanisms utilized during school closures. In addition, this project will develop the capacity of countries to monitor learning after the crisis.
	Early Childhood Development (ECD)	The World Bank team has developed a suite of tools to measure childhood development and early learning quality, including the Anchor Items for Measurement of Early Childhood Development (AIM-ECD), a core set of items with robust psychometric properties across contexts for measuring preschoolers' early literacy, early numeracy, executive functioning, and socioemotional development; Teach ECE, an observation tool that captures the quality of teacher-child interactions in preschools (see above); and the ECD COVID-19 Phone Survey to support countries in capturing the impacts of the pandemic on young children and their families.
Health care	Service Delivery Indicators (SDI) Health	SDI Health provides a set of metrics for benchmarking service delivery performance in primary health care. The overall objective of the indicators is to gauge the quality of service delivery in basic health services measured at the health facility. The indicators enable the identification of gaps and the tracking of progress over time and across countries. It is envisaged that broad availability, high public awareness, and a persistent focus on the indicators will mobilize policy makers, citizens, service providers, donors, and other stakeholders for action to improve the quality of services and, ultimately, to improve development outcomes and social welfare.
	Primary Health Care Performance Initiative (PHCPI)	The PHCPI developed a conceptual framework that describes important components of a strong primary health care system. It is intended to guide what should be measured to inform and drive efforts to improve primary health care. The framework is based on evidence about the key characteristics and determinants of strong primary health care systems, building on existing frameworks for health system performance. The selection of our core indicators and the creation of the PHC Vital Signs Profiles were informed by this framework. The data collected through SDI Health Surveys can be used to help create the PHC Vital Signs Profiles.

(continues on next page)

TABLE 29.1 Survey of Global Initiatives in Education and Health Care Delivery Indicators *(continued)*

Public service	Initiative	Description
Health care *(continued)*	Service Availability and Readiness Assessment (SARA)	SARA is a health facility assessment tool designed to assess and monitor service availability and the readiness of the health sector and to generate evidence to support the planning and managing of a health system. SARA is designed as a systematic survey to generate a set of tracer indicators of service availability and readiness. The survey's objective is to generate reliable and regular information on service delivery (such as the availability of key human and infrastructure resources); on the availability of basic equipment, basic amenities, essential medicines, and diagnostic capacities; and on the readiness of health facilities to provide basic health care interventions related to family planning, child health services, basic and comprehensive emergency obstetric care, human immunodeficiency virus (HIV), tuberculosis, malaria, and noncommunicable diseases.
	Service Provision Assessment (SPA)	The SPA survey is a health facility assessment that provides a comprehensive overview of a country's health service delivery. The SPA looks at the availability of health services; the extent to which facilities are ready to provide health services (do they have the necessary infrastructure, resources, and support systems?); the extent to which service delivery processes meet accepted standards for quality of care; and the satisfaction of clients with the service delivery environment.

Source: Original table for this publication.

These different approaches to measuring service delivery propose a conceptual framework for service delivery and how it should be measured. The common pillars of these conceptual frameworks are the following policy objectives:

- Set priorities regarding improvements in service delivery.

- Identify strengths and gaps in delivery system performance.

- Identify knowledge gaps, where deeper diagnostics are needed.

- Monitor progress on the quality of services.

These policy objectives should guide practitioners in defining the relevant dimensions of quality they are interested in measuring. For example, a practitioner may prioritize improving the quality of student learning in a school. One potential indicator is Learning Poverty: the share of children who haven't achieved minimum reading proficiency. This indicator helps identify the strengths and gaps in service delivery performance by providing objective benchmarks with which to compare student learning across schools. Deeper diagnostics may be required: are there particular age groups that are more vulnerable to low reading proficiency? Are there gender gaps that may be driving these results? Finally, a monitoring strategy allows governments to identify whether progress has been made. For example, the impact on student learning of a policy change such as improving access to school materials can be monitored by taking a baseline and endline survey measuring the indicator.

As practitioners explore these sets of questions, we recommend that they draw upon the international experience of other teams that have developed both conceptual and methodological frameworks to address them. For example, the *World Development Report 2018* (World Bank 2018) provides an array of tools that focus on measuring student learning deficits across the world, and *The Lancet Global Health* Commission on High-Quality Health Systems in the SDG Era (Kruk et al. 2018) provides guidance on how measurement efforts can improve the quality of health care services.[6] As the *World Development Report 2018* argues, measurement makes visible otherwise "invisible" quality deficits in the delivery of educational services. However, while objective measures are necessary, they are not in themselves sufficient to improve the quality of public services. These indicators "must facilitate action, be adapted to country needs, and consist of a range of tools to meet the needs of the system" (World Bank 2018, 91).

This section provides an overview of how MSDs are conceptualized and measured. We divide our discussion into education and health care, corresponding to two distinct but analogous approaches to measuring the quality of public service delivery. We draw on the experience of teams at the World Bank who have

developed and implemented MSDs, as well as other global efforts that have promoted the use of objective measures to improve the quality of service delivery. We present these conceptual frameworks as concrete examples for practitioners interested in developing their own MSDs. Practitioners should bear in mind that these frameworks are neither exhaustive nor prescriptive. We draw from two programs, the SDI Health Survey and the GEPD, to explain what these conceptual frameworks are, why they came to be, and how practitioners may draw on them to develop their own MSDs.

SDI Health Survey Conceptual Framework

The SDI Health Survey is a nationally representative, health facility–based survey that measures the quality of delivery of primary health care services as experienced by citizens across the world.[7] Since its inception in 2008, the objective of the SDI Health Survey has been to improve the monitoring of service delivery to increase public accountability and good governance, as well as targeted interventions through objective measurement of the quality and performance of health services. SDI Health Surveys have been completed in several countries in Africa, including Kenya, Madagascar, and Mozambique (see figure 29.1), and the survey has recently expanded to South Asia (Bhutan), Europe and Central Asia (Moldova), and the Middle East and North Africa (Iraq).

To accomplish this objective, the SDI Health Survey team originally developed a conceptual framework that allowed practitioners to measure the quality of health service delivery. As outlined in Gatti et al. (2021), the first generation of SDI Health Survey content was based on three dimensions, with corresponding topics and associated indicators, as illustrated in table 29.2.

Recently, there has been a push to reimagine how the SDI Health Survey measures the quality of services. In particular, the conceptual framework has been expanded to focus on processes of care and person-centered outcomes (such as patients' experience, including wait time and expenditures incurred). Additional measures have been included to measure job satisfaction and the broader work environment as experienced by health care providers. Finally, given the increasing salience of public health crises, measurements of facilities have been expanded to gauge levels of preparedness for pandemics and disaster scenarios. We provide an overview of the updated conceptual framework and questionnaire modules in figure 29.2.

Both table 29.2 and figure 29.2 present the wide array of health service delivery measurement topics available to practitioners. At the same time, they highlight how foundational concepts—facilities, providers, and patients—can provide a basis to measure the quality of service delivery. Practitioners are encouraged to define policy objectives as outlined at the beginning of this section. For example, a practitioner may prioritize increasing equipment and supplies at the facility level. The SDI Health Survey helps identify strengths and gaps in the delivery system. Certain regions may lag behind others in the availability of these inputs, raising questions as to what is driving this limited availability. Deeper diagnostics may suggest a correlation between regions with lower workforce training and lower equipment availability. The impact of a policy intervention to increase the capacity building of staff may be measured by follow-up SDI Health Surveys.

The Global Education Policy Dashboard Conceptual Framework

The GEPD builds on a set of nationally representative surveys that measure the quality of educational services and learning outcomes, including the SDI Education Survey.[8] Since its initial development in 2019, the GEPD has outlined its goal to measure and highlight the key drivers of learning outcomes, connecting all parts of the production chain. It provides a systemic overview of the drivers of learning, focusing on key dimensions of the educational system, such as teachers and the policies overseeing them. GEPD projects have been completed in Rwanda, Jordan, and Peru, with ongoing implementation in other countries in Africa (Ethiopia and Mozambique), as well as plans for expansion into other regions.

FIGURE 29.1   MSD Health Indicators for a Selection of Countries in Africa

Source: Screenshot of Service Delivery Indicators Health, Interactive Results dashboard, https://www.sdindicators.org/.
*Note:* MSD = measures of service delivery.

The GEPD identifies three key dimensions driving learning outcomes: practices (or service delivery), policies, and politics (figure 29.3). The practices dimension is further divided into four topics: teachers, learners, school inputs, and school management. Because the GEPD also provides indicators for learning, it offers a holistic view of the educational system, connecting outcomes (learning) to their drivers. A total of 39 indicators have been developed to measure these different dimensions. Examples of indicators, as well as their associated topics and dimensions, are provided in table 29.3.[9]

TABLE 29.2 Indicators in the First Service Delivery Indicators Health Surveys

| Dimension	Topic	Indicators
Provider effort	Provider absence	Share of a maximum of 10 randomly selected providers absent from the facility during an unannounced visit
	Caseload per health provider	Number of outpatient visits per clinician per day
Provider's knowledge and ability	Diagnostic accuracy	Percentage of correct diagnoses provided in a selection of five to six clinical vignettes
	Treatment accuracy	Percentage of correct treatments provided in a selection of five to six clinical vignettes
	Management of maternal and neonatal complications	Number of relevant treatment actions proposed by the clinician
Inputs	Medicine availability	Percentage of 14 basic medicines that were available and in stock at the time of the survey
	Equipment availability	Availability and functioning of a thermometer, stethoscope, sphygmomanometer, and weighing scale
	Infrastructure availability	Availability and functioning of an improved water source, an improved toilet, and electricity

Source: Adapted from Gatti et al. 2021.

FIGURE 29.2 Updated Conceptual Framework and Questionnaire Modules for the Service Delivery Indicators Health Survey
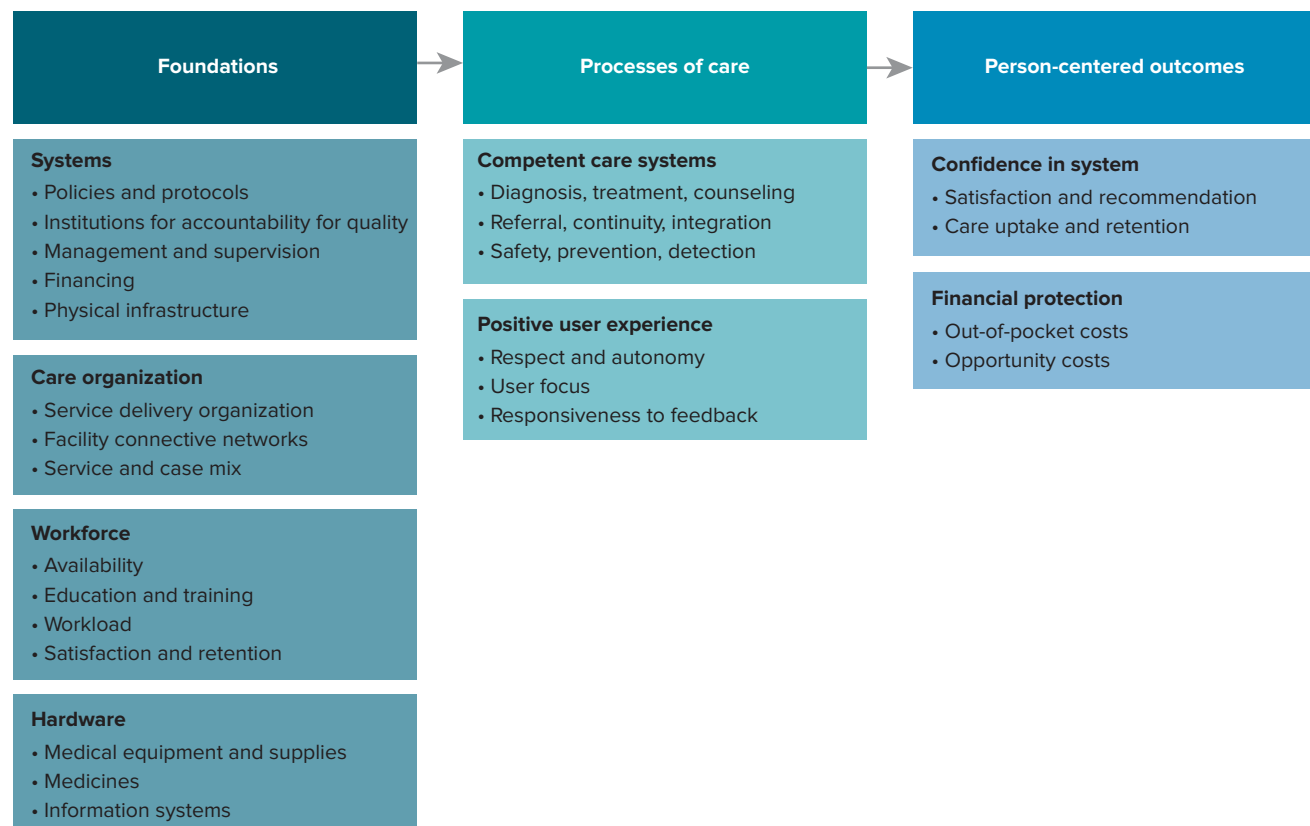
Source: Original figure for this publication.

FIGURE 29.3   Dimensions of the Global Education Policy Dashboard
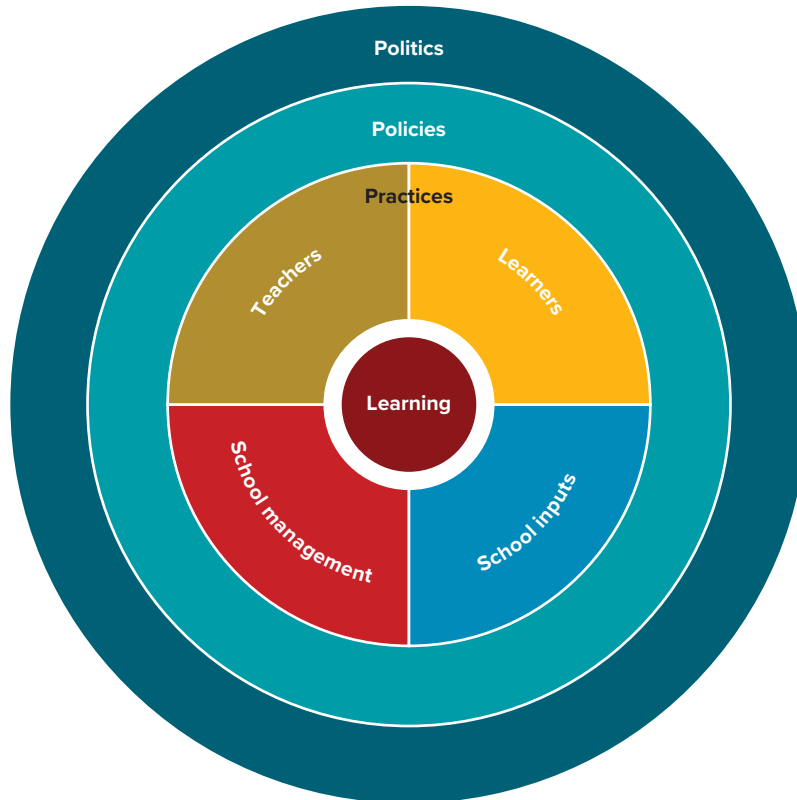
Source: Global Education Policy Dashboard website, https://www.educationpolicydashboard.org/.

We provide an example of how to apply the GEPD indicators to achieve policy objectives. A practitioner may prioritize improving the attraction of teachers. Survey data suggest that while respondents in general perceive recruitment of teachers as meritocratic, the financial incentives of positions are viewed less favorably. A deeper diagnostic shows that these negative perceptions of financial incentives are concentrated among respondents in urban areas, where exit options may be more attractive. As a result, policy makers propose an additional financial bonus for teachers in competitive urban areas. Follow-up surveys monitor the impact of this policy change on the teacher attraction indicator.

As can be seen from the SDI and GEPD examples, contemporary conceptual frameworks provide a holistic assessment of public service delivery. These frameworks include indicators on multiple dimensions, such as workforce, management practices, and welfare outcomes or user experience. As a result, MSDs provide practitioners with granular information on different parts of the delivery chain, enabling targeted interventions. Additionally, granularity allows practitioners to explore causal relationships between dimensions—for example, how worker satisfaction impacts student learning. A holistic assessment of public service delivery therefore serves an important role in better understanding and improving the quality of public service.

PRODUCING MEASURES OF SERVICE DELIVERY

Equipped with a conceptual framework, the next step for practitioners is the actual measurement of service delivery. Within the scope of this chapter, our primary focus is on facility surveys, although the GEPD includes measurements of additional factors, including politics and policies, which are measured through

TABLE 29.3 Global Education Policy Dashboard Indicators

Dimension	Topic	Indicators
Learning	Proficiency on GEPD assessment	Each question on the fourth-grade student assessment is scored as one point. The indicator reports the fraction of students scoring at least 20 out of 24 points on the fourth-grade language assessment and at least 14 out of 17 points on the math assessment.
	Learning Poverty indicator	The Learning Poverty indicator, as reported in the Learning Poverty Database, captures schooling and learning at the end of primary school.
	Teacher content knowledge	This indicator measures the percentage of teachers scoring at least 80 percent correct on the teacher assessment. In this assessment, each question is worth one point.
	Instructional leadership	A score of one to five is assigned based on the presence of four practices as reported by teachers. The four practices, which are given equal weight, are the following: • Had a classroom observation in past year • Had a discussion based on that observation that lasted longer than 30 minutes • Received actionable feedback from that observation • Had a lesson plan and discussed it with another person.
Policies	Teaching—attraction	A score of one to five is assigned based on five factors. Each factor receives an equal weight in terms of possible points (0.8). The factors are the following: • Job satisfaction • Community satisfaction • Perceived meritocracy • Financial incentives • Absence of salary delays.
	School management—evaluation	A score of one to five is assigned based on four factors. Each factor receives an equal weight in terms of points. The factors are the following: • Reported evaluation in the past year (1) • Reported multiple evaluation criteria (1) • Reported consequences for negative evaluation (1) • Reported consequences for positive evaluation (1).
Politics	Characteristics of bureaucracy	A score of one to five is assigned based on four factors. Each factor has been given an equal weight. Each factor is based on a set of three to four questions, each scored one to five. For each factor, the average score across the questions is determined. To construct the total score, the average is taken of the four factor scores. The factors include the following: • Knowledge and skills • Work environment • Merit • Motivation.
	Impartial decision-making	A score of one to five is assigned based on four factors. Each factor has been given an equal weight. Each factor is based on a set of three questions scored one to five. For each factor, the average score across the three questions is determined. To construct the total score, the average is taken of the four factor scores. The factors include the following: • Politicized personnel management • Politicized policy making • Politicized policy implementation • Employee unions as facilitators.

Source: Adapted from GEPD 2021b.

surveys of public officials (discussed in detail in part three of *The Government Analytics Handbook*). Facility surveys are only one of many different, important ways of measuring service delivery. Many things are not measured by visiting facilities: children who do not attend schools are not included, clients who do not visit clinics are not included, and central-governance-level issues, such as national policies or district-level protocols, are not measured. In this section, we provide guidelines on how practitioners can move forward and generate their own MSDs at the facility level.

We divide the production of facility-based MSDs into four stages: design, implementation, analysis, and dissemination. This section is filled with practical advice on how our teams have engaged in the rollout of MSDs (we have drawn especially on GEPD [2021a] and SDI [2019]). Practitioners are invited to adapt these implementation guidelines to their own contexts and needs.

Design: Stakeholder Engagement and Survey Instrument

The first step is securing engagement and institutional support from MSD stakeholders. Depending on the target public service, these stakeholders may vary. Once the indicators are produced, who will consume these data? Where in public administration would these indicators have a maximal impact? These questions should help practitioners identify relevant actors. For education, stakeholders may include the ministry of education, subnational governments responsible for providing these services, and civil society organizations, such as teacher associations. For health care, potential stakeholders would be the ministry of health, development partners (such as the WHO and UNICEF), and heads of clinics and hospitals, among others. Stakeholder engagement should guide the selection of relevant dimensions and indicators: what specific actions are expected to change because of the MSDs?

Practitioners are encouraged to assess trade-offs: broader coverage in terms of dimensions can come at the expense of depth in particular dimensions—such as teacher skills—that may be of greater interest to stakeholders. Perhaps resources are constrained, and engaging in a full-fledged survey is not feasible. If this is the case, practitioners may have to select a few dimensions that are considered priorities by stakeholders. However, this selectiveness implies a cost: the SDI Health Survey and the GEPD dimensions speak to one another and allow for data triangulation to get a comprehensive picture of the service across key domains. The selection of dimensions may also compromise international benchmarking for mutual learning, which requires comparability between and within countries.

Finally, practitioners should define the level of representativeness of the survey. In some cases, a nationally representative survey will suffice, reducing the burden and cost of implementation. Such statistics are useful for national policy makers to formulate broad changes to the service delivery system as a whole. For example, if there are systemic issues in the distribution of facility inputs, such as schoolbooks, national statistics allow for a broad response. In other contexts, subnational variation in the quality of service may be of interest and is a powerful complement to national statistics. Often, the problems facing subnational regions differ, with some regions requiring facility inputs and others requiring staff training. Gaining this greater degree of granularity requires a different sampling strategy. For example, the GEPD strategy follows stratified random sampling, defining strata as subnational regions and ensuring that all relevant geographical divisions are included. The "GEPD Technical Note" (GEPD 2021b) covers other specifications, such as survey weights and power calculations, as well as data collection and quality checks.

Implementation: Training and Data Collection

Once the design phase is completed, the production cycle moves to implementation. Both SDI and GEPD teams use in-person surveys to undertake their assessments. There are important benefits to practitioners interested in measuring service delivery quality at the facility level in directly collecting data through field surveys, rather than relying on administrative data alone. Administrative data are often unreliable, in particular for areas of the country in which information systems are not widely available, as is often the case in rural facilities. Additionally, administrative data may be subject to misreporting, whereas enumerators serve as third-party observers. Finally, surveys can complement the development of robust information systems, providing actionable data for a fraction of the time and cost.[10] Data collection may either be done in-house or through public procurement of a survey firm.

Note that only if the accuracy of the data is guaranteed can practitioners generate robust analytical insights through indicators.[11] Otherwise, indicators will reproduce the biases and inaccuracies of the data, providing stakeholders with faulty evidence for policy making. Data accuracy relies on a robust, well-adapted, and piloted survey tool; high-frequency and sense checks; and data validation procedures, such as revisits or callbacks, to samples of the same facilities. These data validation procedures include verifying the time of submission of survey responses, the length of interviews, and systematic missingness in variables, among other checks.[12]

Both the GEPD and the SDI Health Survey engage in the automation of data collection, which is enabled by the use of Survey Solutions, an open-source tool available free of charge.[13] This has improved data quality over the past decade, as electronic data collection technologies have improved and become more pervasive. These technologies make real-time monitoring of data quality—through high-frequency checks and data quality warnings—easier to implement. Chapter 5 of the *Handbook* outlines a variety of protocols to ensure data quality, such as enumerator training and high-frequency checks on every batch of data, typically every day.

One innovation that both the SDI Health team and the GEPD leverage is the combination of announced and unannounced visits.[14] The former allows for a more thorough discussion of topics for which the service delivery providers need to prepare materials and information. Unannounced visits, on the other hand, enable practitioners to identify behaviors or practices that frontline providers or facility managers may have an incentive to either conceal or misreport, as well as those that may be disrupted by conducting a survey. The goal is not to reprimand providers or identify evidence of misconduct but rather to provide a more accurate assessment of the practices that occur during an average day of public service delivery. For example, one of the indicators collected by the SDI Health and Education Surveys is the health care or education provider's presence or absence during an unannounced visit.

Analysis: Data Validation and Production of Indicators

Once the data have been collected by enumerators, the next step is validating the data, cleaning them, analyzing them, and generating indicators. Data validation should ideally be conducted in an automated fashion, where checks are encoded into relevant software and thus replicable in other settings. For example, the GEPD leverages open-source statistical software called R to validate the data collected in a documented and replicable way.[15] Data validation and processing are often challenging, in particular when in-house data analysis capacity may be more limited. In these contexts, we encourage practitioners to access different resources that document best practices in survey analysis, such as DIME Analytics (Bjärkefur et al. 2021). Chapter 5 of the *Handbook* provides additional information on this program.

Once data have been thoroughly validated and cleaned, the indicators can be generated. Again, where possible, these indicators should be generated through replicable steps, preferably using open-source software.[16] However, other statistical software, such as Microsoft Excel, may be capable of generating the relevant indicators. The crucial step is that the process for generating the indicators be documented, transparent, and replicable.

Dissemination: Stakeholder Presentations and Wider Communication Efforts

The final step in the production cycle for MSDs is the dissemination of the results to stakeholders, as well as wider communication efforts. In general, we have found that interactive dashboards are an important component of MSD dissemination. Dashboards allow for intuitive visualization of the different dimensions of public service and empower stakeholders to interact with the data at a greater level of granularity than static reports. For an example, see figure 29.4.[17] Practitioners may click on different dimensions and obtain additional information for particular indicators. Colors allow users to identify where in the public service delivery chain more attention is needed.

An important feature of the GEPD is that it goes beyond the measurement of the indicators, providing visual feedback on each indicator to guide policy making at a granular level.[18] The feedback comprises three colors: red (needs improvement), yellow (caution needed), and green (on target). This visual feedback allows policy makers to immediately identify topics in which additional work is needed, as well as policy areas in which targets have been met (figure 29.4). This action-oriented visualization allows practitioners to design their educational policy with an intuitive and evidence-based approach.

This systematic view, with diagnostics associated with each topic, also allows practitioners to hone in on specific areas that require further development. For instance, in figure 29.4, we find that Peru has done quite

FIGURE 29.4   Example of Global Education Policy Dashboard

Peru ⌄

To understand learning and participation outcomes, it is important to understand what is happening inside the schools. Here you can find the outcome indicators capturing learning for all (that is, learning combined with access) at the center, ringed by indicators representing the four-main school-level service-delivery factors: teaching, school management, inputs & infrastructure, and prepared learners. Click through the indicators to learn more about them as well as the indicators of deeper systematic characteristics linked with each of them.

INPUTS AND INFRASTRUCTURE

Basic Inputs

Basic Infrastructure

TEACHING

Teacher Presence

Content Knowledge

Pedagogical Skills

OUTCOMES

Low learning and lack of access to school can leave children without foundational skills. These two sets of outcomes make up the learning poverty rate.

Learning Poverty Rate

LEARNERS

Readiness for Learning

Student Attendance

SCHOOL MANAGEMENT

Operational Management

Instructional Leadership

School Knowledge

Management Practices

✓ On Target   ! Caution   ✕ Needs Improvement

Source: Screenshot of GEPD dashboard, https://www.educationpolicydashboard.org/practice-indicators/per.

well in management practices at the school level. However, teachers' skills require more attention: in particular, pedagogical skills and content knowledge. Thus, practitioners can prioritize certain areas over others, as they gather granular and actionable evidence on how indicators are faring.

While dashboards may be helpful for consumption within the government, a wider audience can be reached by organizing events where the MSDs are revealed to the public. The main findings of the MSDs should be presented by stakeholders and the implementation team, fostering accountability and transparency. Additionally, facilities that were interviewed could be given both the survey results and actionable steps they can take to improve their indicators, as well as recognition for areas in which they are successful. As noted in chapter 26 of the *Handbook*, governments should move beyond survey indicators by providing immediate feedback to facilities and civil servants.

A GLOBAL EFFORT TO IMPROVE PUBLIC SERVICES

So far, this chapter has outlined some key considerations in measuring public sector service delivery in primary health care and education, as described by experts in the SDI Health and GEPD teams. These teams are part of a wider community of practice on generating measures to improve public service delivery, and we

encourage practitioners to connect to other global efforts. By engaging with this global community, governments can benefit from knowledge exchange with international organizations, as well as other practitioners pursuing similar initiatives.

Harmonizing surveys and producing indicators that can be benchmarked to other countries provides practitioners with objective standards against which to measure themselves—to help them understand, for instance, whether an enrollment rate is particularly high or low. If multiple countries have fielded similar surveys, a country team can take an indicator and compare it to other countries with similar educational systems and levels of economic development. Additionally, global indicators on public service delivery provide a public good that can be shared and accessed by communities of practitioners across the world.

The SDI and the GEPD are key players in the global movement to generate MSDs, with tools and expertise to help in this endeavor. Connecting to these global initiatives enables practitioners to capitalize on decades of experience, tools, and technical expertise that teams (like those in the World Bank) can offer to help optimize the long-lasting impact of these endeavors. An important exercise that global engagement enables is international benchmarking, which helps practitioners understand how well their countries are performing relative to others. Benchmarking exercises allow countries to quantify how far they may be from the frontier and to learn from the best in class what they can do to improve it.

This global community also makes available teams of education and health service delivery experts that can guide practitioners through the implementation of MSDs and accelerate rollout. Located in the World Bank's Education Global Practice and Health, Nutrition, and Population Global Practice, global experts provide technical assistance to practitioners interested in developing MSDs. While production cycles vary, the entire process from design to dissemination generally takes around one year. The costs vary as well but average US$100,000–US$400,000, based on country experiences.

Practitioners are encouraged to reach out to the SDI Health team for further details.[19] The materials and services provided by the SDI Health team include the following:

- Standardized health facility survey materials, field manuals, training materials, and suggestions for adaptation of the survey instrument

- Technical guidance on survey design and sampling strategy

- Assistance with quality control during data collection

- Capacity building for governments to generate and disseminate MSDs.

The GEPD provides similar services. It builds on the MSD framework but also leverages insights from other initiatives within the World Bank, such as the Systems Approach for Better Education Results (SABER) (World Bank 2020) and the Global Survey of Public Servants (GSPS).[20] Both the SDI Health Survey and the GEPD share a focus on capacity building, providing practitioners with the tools and resources necessary to reproduce conceptual and methodological frameworks on the ground. This approach ensures the co-ownership of results and operational relevance.

CONCLUSION

This chapter has argued that MSDs provide governments with tools to measure and improve the quality of public service delivery. MSDs increase the accountability of governments because stakeholders gain access to objective measures of how public services are operating. Both the GEPD and the SDI Health Survey can provide governments with a systematic overview of service delivery, unpacking welfare outcomes—student learning and health care—as well as the different chains of delivery. We have also presented a step-by-step

guide to producing MSDs, drawing on the experience of teams at the SDI Health Survey and GEPD programs at the World Bank.

However, measurement alone is not enough to improve the quality of public services. MSDs need to be linked directly to stakeholders who can enact change in the delivery system. The broader public has to be made aware of the results as well. Moreover, as emphasized in chapter 4 of the *Handbook*, measurement is not a substitute for the proper management of services. With these caveats in mind, MSDs can allow practitioners and the broader research community to better understand the drivers of health and education outcomes. These efforts are part of a global agenda: we encourage readers to learn more from the publicly available resources listed here and, if interested, to reach out to relevant expert teams.[21]

Improving the quality of service delivery is a complex endeavor. As the COVID-19 pandemic has high-lighted, unexpected crises can have profound consequences for the quality of health care and educational services. MSDs should be responsive to these sudden changes, as well as more gradual, evolving needs. Rather than ends in themselves, indicators should be used as tools to improve what ultimately matters: the lives of citizens who rely on public services.

NOTES

1. This chapter uses the neutral term *absence* rather than *absenteeism*; the focus in the surveys is on the effect of provider absence on the quality of service delivery rather than on assigning blame to providers, who may be absent for reasons that are out of their control.
2. Primary health care, which was enshrined in the World Health Organization's Alma-Ata Declaration, includes essential services in health care, such as prenatal care and basic diagnostics. Primary education refers to pre-secondary education, including primary and middle school.
3. For an overview of these different dimensions in primary health care, see Andrews and Sharma (2021).
4. Additionally, there are often cases in which citizens still lack access to the basic services of health care and education. As a result, any improvements in these services may fail to improve their lives. Ensuring that broad access to these services develops in parallel with improvements in their quality is therefore crucial.
5. Further information about the PHCPI is available at the initiative's website, https://improvingphc.org/.
6. We highly recommend chapter 4 of the *World Development Report 2018*, "To Take Learning Seriously, Start by Measuring It."
7. To learn more about this program, see the SDI website, https://www.sdindicators.org/.
8. The GEPD School Survey builds on the following surveys: the Service Delivery Indicators (SDI) Survey, on teachers and inputs/infrastructure; Teach, on pedagogical practice; the Global Early Child Development Database (GECDD) and the Measuring Early Learning Quality and Outcomes (MELQO) initiative, on the school readiness of young children; and the Development World Management Survey (DWMS), on management quality. The GEPD also includes data on public officials from the Survey of Public Servants. For additional details, see the GEPD website, https://www.educationpolicydashboard.org/.
9. For a complete presentation and discussion of the indicators, see GEPD (2021b).
10. For a broader discussion of information systems, refer to chapter 9.
11. This is a similar argument to the one presented in chapter 9.
12. Examples of these checks for the GEPD can be found in the RMD file "School Data Quality Checks," located in the School folder in the Master Code directory in the GEPD repository in the World Bank's GitHub repository, available at https://github.com/worldbank/GEPD/blob/master/Master_Code/School/school_data_quality_checks.Rmd (latest commit September 24, 2019).
13. For more information, see the Survey Solutions website, https://mysurvey.solutions/en/.
14. The GEPD team provides a two-week window in which the visit will take place but does not disclose the precise date.
15. R is freely available at https://cran.r-project.org/. The entire GEPD repository and code is available in the World Bank's GitHub repository at https://github.com/worldbank/GEPD (latest commit February 23, 2023).
16. The GEPD also makes available the code generating the indicators in the Indicators folder in the GEPD repository in the World Bank's GitHub repository, available at https://github.com/worldbank/GEPD/tree/master/Indicators (latest commit January 13, 2023).

17. The dashboard is available on the GEPD website at https://www.educationpolicydashboard.org/practice-indicators/per.
18. This is similar to the visual feedback provided by the Employee Viewpoint Survey Analysis and Results Tool (EVS ART), a dashboard described in chapter 9, case study 9.3, and in chapter 26 of the *Handbook*.
19. The SDI Health team can be contacted at sdi@worldbank.org.
20. For more information on the Global Survey of Public Servants (GSPS), see its website, https://www.globalsurveyofpublicservants.org/.
21. The GEPD team can be contacted at educationdashboard@worldbank.org. The SDI Health team can be contacted at sdi@worldbank.org.

REFERENCES

Amin, Samia, Jishnu Das, and Markus Goldstein, eds. 2007. *Are You Being Served? New Tools for Measuring Service Delivery*. Washington, DC: World Bank.

Andrews, Kathryn, Ruben Conner, Roberta Gatti, and Jigyasa Sharma. 2021. "The Realities of Primary Care: Variation in Quality of Care across Nine Countries in Sub-Saharan Africa." Policy Research Working Paper 9607, World Bank, Washington, DC.

Andrews, Kathryn, and Jigyasa Sharma. 2021. "A Revolution in Health Service Delivery Measurement." *Investing in Health* (blog). *World Bank Blogs*, July 1, 2021. https://blogs.worldbank.org/health/revolution-health-service-delivery-measurement.

Bjärkefur, Kristoffer, Luíza Cardoso de Andrade, Benjamin Daniels, and Maria Ruth Jones. 2021. *Development Research in Practice: The DIME Analytics Data Handbook*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/35594.

Gatti, Roberta, Kathryn Andrews, Ciro Avitabile, Ruben Conner, Jigyasa Sharma, and Andres Yi Chang. 2021. *The Quality of Health and Education Systems across Africa: Evidence from a Decade of Service Delivery Indicators Surveys*. Washington, DC: World Bank.

GEPD (Global Education Policy Dashboard). 2019. "School Data Quality Checks." School, Master Code, GEPD, World Bank. GitHub repository, September 24, 2019. https://github.com/worldbank/GEPD/blob/master/Master_Code/School/school_data_quality_checks.Rmd.

GEPD (Global Education Policy Dashboard). 2021a. "Global Education Policy Dashboard Implementation Brief." Global Education Policy Dashboard, World Bank Group. https://www.educationpolicydashboard.org/sites/epd/files/resources-documents/GEPD%20Implementation%20Brief.pdf.

GEPD (Global Education Policy Dashboard). 2021b. "GEPD Technical Note." Global Education Policy Dashboard, World Bank Group, April 19, 2021. https://www.educationpolicydashboard.org/sites/epd/files/resources-documents/GEPD%20Technical%20Note.pdf.

Hanson, Kara, Nouria Brikci, Darius Erlangga, Abebe Alebachew, Manuela De Allegri, Dina Balabanova, Mark Blecher, Cheryl Cashin, et al. 2022. "*The Lancet Global Health* Commission on Financing Primary Health Care: Putting People at the Centre." *The Lancet Global Health Commission* 10 (5): E715–E772. https://doi.org/10.1016/S2214-109X(22)00005-5.

Kruk, Margaret E., Anna D. Gage, Catherine Arsenault, Keely Jordan, Hannah H. Leslie, Sanam Roder-DeWan, Olusoji Adeyi, et al. 2018. "High-Quality Health Systems in the Sustainable Development Goals Era: Time for a Revolution." *The Lancet Global Health Commission* 6 (11): E1196–E1252. https://doi.org/10.1016/S2214-109X(18)30386-3.

SDI (Service Delivery Indicators). 2019. "Nuts and Bolts: A Brief Guide for Task Teams." Unpublished document, September 24, 2019, World Bank, Washington, DC.

WHO (World Health Organization). 2010. *Monitoring the Building Blocks of Health Systems: A Handbook of Indicators and Their Measurement Strategies*. Geneva: WHO Press. https://apps.who.int/iris/bitstream/handle/10665/258734/9789241564052-eng.pdf.

World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank. https://www.worldbank.org/en/publication/wdr2018.

World Bank. 2020. "Systems Approach for Better Education Results (SABER)." Education Brief, World Bank. https://www.worldbank.org/en/topic/education/brief/systems-approach-for-better-education-results-saber.

World Bank. 2021. "What Is Learning Poverty?" Education Brief, World Bank, April 28, 2021. https://www.worldbank.org/en/topic/education/brief/what-is-learning-poverty.