The Quality of Health and Education Systems Across Africa

Evidence from a Decade of Service Delivery Indicators Surveys

Roberta Gatti
Kathryn Andrews
Ciro Avitabile
Ruben Conner
Jigyasa Sharma
Andres Yi Chang

WORLD BANK GROUP

The Quality of Health and Education Systems Across Africa

The Quality of Health and Education Systems Across Africa

Evidence from a Decade of Service Delivery Indicators Surveys

Roberta Gatti, Kathryn Andrews,
Ciro Avitabile, Ruben Conner,
Jigyasa Sharma, and Andres Yi Chang

WORLD BANK GROUP

Contents

Boxes

Figures

Tables

Acknowledgments

Reference

Andrews, K., R. Conner, R. Gatti, and J. 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.

About the Authors

Roberta Gatti is the World Bank's chief economist of the Middle East and North Africa region and former chief economist of the Human Development Practice Group, where she led the SDI and the Human Capital Index initiatives. She joined the World Bank in 1998 as a Young Professional in the Development Research Group. Her research includes theoretical and empirical contributions to labor and household economics, political economy, growth, and social inclusion. She has authored multiple World Bank flagship reports, including *Jobs for Shared Prosperity* and *Being Fair, Faring Better*. She has taught at Georgetown University and Johns Hopkins University and holds a BA from Università Bocconi and a PhD in economics from Harvard University.

Kathryn Andrews is a health economist in the South Asia Health, Nutrition, and Population Global Practice, leading analytical work related to health service delivery measurement and supporting operational work on health systems strengthening. She was previously a Young Professional in the Human Development Chief Economist's Office, where she supported the design and implementation of Service Delivery Indicators (SDI) surveys. She previously worked at the World Health Organization and Institute for Health Metrics and Evaluation, and holds an ScD in Global Health and Population from the Harvard T. H. Chan School of Public Health.

Ciro Avitabile is a senior economist in the Latin America and the Caribbean Education Global Practice, where he works on supporting investments for human capital development based in Lima, Peru. Before joining the World Bank, he was an economist at the Inter-American Development Bank; he holds a PhD in economics from University College London.

Ruben Conner is pursuing his DrPH at Johns Hopkins University. Previously, as an extended consultant in the Human Development Chief Economist's Office, he guided SDI surveys in Armenia, Libya, and Moldova. Before joining the World Bank, he engaged in outbreak response for Doctors Without Borders in South Sudan and worked on malaria programs at PATH. His areas of focus include the cost and efficiency of health systems, large-scale impact evaluations, and vaccination planning.

Jigyasa Sharma is a health economist in the Health, Nutrition, and Population Global Unit, supporting and leading analytical work related to primary health care and health system strengthening and resilience. She was previously a Young Professional in the Human Development Chief Economist's Office. She holds an ScD in global health and population from the Harvard T. H. Chan School of Public Health.

Andres Yi Chang is an economist in the Human Development Chief Economist's Office, where he also serves as the SDI program manager. He began working at the World Bank in 2015, first at the World Bank Research Development Group and then moving to the HDCEO in 2019. Before joining the World Bank, he worked at Yale University, the Organization of American States, the Pan American Health Organization, and Leibniz University of Hannover. His research and projects are related to education, health, and social development in low- and middle-income countries, particularly in Africa, Latin America, and Southeast Asia. He has a BA in economics from University of California, Berkeley, and an MA international and development economics from Yale University.

Abbreviations

DHIS2   District Health Information Software (Kenya)

DHS   Demographic and Health Survey

D-WMS   Development World Management Survey

GDP   gross domestic product

ICT   information and communication technology

IMCI   integrated management of childhood illness

IT   information technology

IV   intravenous

L1   an individual's first language learned ("mother tongue")

mhGAP   Mental Health Gap Action Programme (World Health Organization)

NCD   noncommunicable disease

NG   nasogastric

NGO   nongovernmental organization

OECD   Organisation for Economic Co-operation and Development

ORS   oral rehydration salts

PHC   primary health care

PIRLS   Progress in International Reading Literacy Study

PISA   Program for International Student Assessment

SABER   Systems Approach for Better Education Results

SDI   Service Delivery Indicators

WHO   World Health Organization

Executive Summary

Health and education systems play a key role in human capital formation

When we talk about human capital, we mean the skills and health that people accumulate over their lifetime. Human capital is a cornerstone of prosperity, poverty eradication, and social cohesion. Human capital accumulation is the result of the interaction of many actors, including individuals, families, governments, the private sector, and civil society. Education and health systems have an important role to play in this process, not only to ensure equitable access but also to provide services with high quality. Yet, even before the COVID-19 (coronavirus) pandemic, persistent gaps in health and learning outcomes between and within economies highlighted the urgency of improving the quality of education and health care.

Are systems delivering for citizens?

To get the best results from their investments in health and education, policy makers need evidence on how systems are working for citizens.

Service Delivery Indicators (SDI) surveys supply this evidence by measuring key aspects of the quality of education and health services, collecting data in public and private, rural and urban schools, clinics, and hospitals. The SDI approach is premised on the *World Development Report 2004: Making Services Work for the Poor*, which emphasizes empowering citizens with information for better accountability (World Bank 2004).

SDI surveys focus on the experience of average citizens in primary health care and primary schools. The surveys collect data on the following questions:

- *What do providers know?* SDI surveys directly test teachers' knowledge of the material they are supposed to teach and health care providers' ability to diagnose and treat common medical conditions. They are the only surveys of this scale that measure the knowledge of health care providers comparably across countries.
- *What do providers do?* Visiting unannounced, SDI enumerators assess the level of absence among teachers and health workers. SDI surveys have generated the only internationally comparable data on provider absence in the health and education sectors, quantifying what many had only hypothesized.
- *What tools do providers have?* SDI enumerators directly check the presence of basic infrastructure, medical equipment, and essential medicines in health facilities. In schools, enumerators visually check that children have textbooks, classrooms have blackboards, and each school has toilets separated by gender.
- *What are children learning?* SDI education surveys measure key education outcomes by testing fourth graders' language, math, and nonverbal skills.

A decade of data show that schools and health clinics are still falling short across Africa

Begun in African countries and now expanding to low- and middle-income countries around the world, SDI surveys have generated nationally representative data that enable governments and citizens to identify gaps and benchmark progress within and across countries over time. The book describes how SDI data have been used to shape public debate and policy action in countries across Africa, providing insights on the role of good measurement for reform. Today, as technology transforms data collection in health and education systems, the book also charts an agenda for measuring how countries develop policies, design interventions, and deliver essential services.

Levers for health system change: Strengthen skills and tools on the frontlines

The SDI health surveys offer important lessons for strengthening—and, if necessary, transforming—primary health care systems in the wake of COVID-19. Data from across nine countries in Sub-Saharan Africa in the past decade highlight patterns in service delivery, identifying both the strengths and the fragility of existing primary care systems. Despite decades of investment to improve structural inputs to care, many facilities lack the basic necessities, including essential medicines, simple diagnostic equipment, and adequate water and sanitation infrastructure. The clinical abilities of health care providers, measured as their diagnostic and treatment accuracy on patient case simulations (clinical "vignettes"), vary substantially by country; but providers at lower-level facilities score noticeably worse. This is particularly troubling, given that these facilities are where people typically make first contact with the health system.

The surveys also reveal a high rate of absence from facilities among health care providers, exceeding 20 percent in almost all countries, although the majority of this absence is authorized. These data also put numbers on the much-debated issue of efficient allocation of human resources. Across all facilities, the average health care provider in these countries sees 13 patients per day, varying from a low of 3 patients per provider per day in some facilities in Nigeria to a high of 23 patients per provider per day in Mozambique.

Collectively, these findings bolster the case for key health reforms and specific investments to strengthen service delivery. Absences, which are higher at public facilities, could potentially be reduced with better supervision and management. To better tailor service provision, governments could consider reallocating human resources. Competencies among health care providers at lower-level facilities should be reinforced, with pre- and in-service training focused on the diagnosis and treatment of common health conditions. Shortages of medicines and equipment impede basic care, and further investigation is needed to see why these shortages persist. Finally, infrastructure gaps are particularly pronounced at rural facilities, and governments can continue to close such gaps by improving the supply of water, sanitation, and electricity.

Improving schools: Countries can learn from their own top performers

Education systems are also under duress from COVID-19, with school closures and learning losses exacerbating preexisting inequities. Across the SDI

countries, measured student learning is low, with less than half of students able to read a simple sentence or perform basic mathematical operations. However, differences in learning are large, both between and within countries, with the lowest-performing schools concentrated in rural areas. There is also important variation in the factors associated with learning. Almost a quarter of teachers can be expected to be absent from school during an unannounced SDI visit, and, among those present in school, many are not present in the classroom during their assigned teaching times. Teachers' skills are also low, in both content knowledge and pedagogy, and are not correlated with education level.

Identifying some of the top schools in each country and analyzing how they succeed could help to set examples for lower-performing schools. Basic school inputs, such as whiteboards and functioning toilets, can be effective in improving the schooling experience. Hiring more teachers—and basing that recruitment on talent instead of solely on credentials—would help to lower the high ratio of students per teacher and improve learning outcomes. Finally, private schools often outperform their public counterparts in student learning, and examining the drivers of their success could help to spark innovation in the public sector.

Data to drive change

By measuring how services are delivered, SDI surveys link resources to results and help to shift the national policy dialogue from inputs to quality and outcomes. SDI surveys have provided the necessary evidence to spur policy debates and accelerate reforms. For example, in Mozambique, a national campaign to tackle teacher absence was instituted after the launch of the report of the 2014 SDI education survey (Bassi, Medina, and Nhampossa 2019). In Togo, teachers' college curricula were revised in response to the finding that only 2 percent of fourth-grade teachers scored 80 percent or more on the grade-level test. With the approval and support of Togo's education ministry, a teachers' union delivered regional workshops aiming to improve learning outcomes. In Tanzania, SDI data were used as diagnostics during the planning of major reforms, and indicators from SDI surveys were added to the monitoring framework of the Big Results Now! education program.

Besides informing dialogue and reform, SDI surveys provide a platform for innovation and research. In education, the SDI initiative has recently expanded into a novel teacher observation tool to enrich the understanding of pedagogy. In health, researchers are adapting SDI questionnaires to capture different aspects of quality of care. For example, clinical vignettes have been developed to assess clinicians' knowledge about the nutrition and growth trajectories of children and about common noncommunicable diseases, such as depression and

hypertension. To adapt to the evolving view of health care as a patient-centered service, comprehensive patient exit surveys on experience and satisfaction with care are being piloted. New modules to measure the quality of management and assess health care providers' well-being and work environment can provide more granular information on the levers available to policy makers.

The COVID-19 pandemic has reinforced the need to rethink the delivery of health and education services. At a time when governments face competing demands and fiscal space is tight, more and better measurement of the factors that determine quality in health care and learning can help countries to protect hard-won gains in human capital and secure the foundations of sustained, inclusive growth. The insights learned from SDI surveys and continuous innovation in the SDI measurement platform offer important contributions toward achieving this goal.

References

Bassi, M., O. Medina, and L. Nhampossa. 2019. "Education Service Delivery in Mozambique: A Second Round of the Service Delivery Indicators Survey." World Bank, Washington, DC.

World Bank. 2004. *World Development Report 2004: Making Services Work for the Poor.* Washington, DC: World Bank.

1. Introduction

Human capital at the core of development

Human capital—the skills and health that people accumulate over their lifetime—is a cornerstone of prosperity, poverty eradication, and social cohesion. In turn, strong health and education systems are necessary for human capital accumulation and lay the groundwork for countries' long-term economic success (Flabbi and Gatti 2018; World Bank 2020a).

Even before the COVID-19 (coronavirus) pandemic, many countries were struggling with low human capital outcomes. The World Bank's 2020 Human Capital Index update shows that, globally, a typical child born just before the pandemic could expect to reach only 56 percent of her productivity as a future worker (World Bank 2020a). This statistic reflects decades of progress in extending life expectancy and expanding schooling access and quality, but it also points to the long road that lies ahead. Persistent gaps in health and learning outcomes between and within countries highlight the need to provide equitable access to health and education opportunities for all. Investing in the health and education pillars of human capital today, in the countries where needs are greatest, will bring both immediate and longer-term benefits (Kim 2018). Failure to make these timely investments will make it harder and costlier to reduce inequalities in the future.

COVID-19: Challenging the resilience of health and education systems

Today, COVID-19 poses new threats to human capital accumulation. The pandemic has dramatically affected health, not only through its toll on human lives but also through potential long-term reductions in healthy functioning among survivors of the disease (Mahase 2020). School closures and the shift to remote learning have reduced access to education and may prove particularly damaging for children from disadvantaged backgrounds and for girls (World Bank 2020b). Along with these direct impacts, the pandemic is also affecting systems and people's lives through indirect mechanisms, including disruptions in non-COVID-19 essential health services and sharp declines in family income following lockdowns (Corral and Gatti 2020).

The pandemic has affected both the supply of and demand for essential care, as providers have struggled to keep up with the care needs of patients with COVID-19; at the same time, many citizens have chosen to avoid health facilities, often forgoing needed care, in order to avoid potential COVID-19 exposure. Experience from previous epidemics suggests that the indirect effects on population health may ultimately cause greater harm than the virus itself (Elston et al. 2017). Evidence already indicates sharp drops in routine immunization coverage during COVID-19 (Jain et al. 2020). The impact on maternal and child health is expected to be especially severe. Early model predictions at the global level suggest a monthly increase in maternal and child mortality of up to 39 percent and 45 percent, respectively, attributable to COVID-19–related poor nutrition and interruption of essential health services (Roberton et al. 2020).

On the education side, there is a potential "lost generation" of learners, reflecting the direct threat of the crisis, the prolonged interruption of schooling, and expected increases in poverty (UNICEF 2020). Worldwide school closures due to COVID-19 are likely to worsen learning disparities, as disadvantaged families are less able to facilitate home-based learning for their children. Dropout rates have increased, and many students no longer benefit from nutritious school meals or the structure, social engagement, and general support afforded by schools (World Bank 2020b).

Although data on the pandemic's impact on human capital are just trickling in, simulations show that the COVID-19 crisis may roll back a decade of hard-won human capital gains unless significant investments are made to protect human capital. With 26 million to 40 million additional people pushed into extreme poverty in 2020 (World Bank 2020c), disruptions in immunizations and other routine health services (WHO 2020), and widespread school closures (Human Rights Watch 2020), the COVID-19-induced risks are especially critical in Africa. The global economic downturn may also result in declining levels

of public revenue and rising levels of debt, which will increase fiscal constraints and make using resources wisely even more imperative.

Leveraging the crisis response to improve systems

Given these challenges, building the strength and resilience of health and education systems has taken on new urgency in Africa and elsewhere. Many governments have already begun building flexibility and adaptability into systems, both to provide services during the emergency and to ensure that systems are better able to respond to changing environments in the future.

Resilience has been defined as "the ability of people, households, communities, countries, and systems to mitigate, adapt to, and recover from shocks and stresses in a manner that reduces chronic vulnerability and facilitates inclusive growth" (USAID 2012). The current pandemic has shown variability in the adaptiveness of health systems. Many health facilities were underprepared for the initial wave of COVID-19 cases, with very limited capacity to provide critical care and little guidance on isolation and containment procedures. In addition, many lower-level facilities lacked adequate handwashing facilities, had little or no personal protective equipment, and did not have the necessary communication equipment (Sharma et al. 2020). Similarly, education systems have been strained by the abrupt shift to remote learning and have had to make quick decisions on when and how schools should remain open.

The current crisis highlights the centrality of health and education systems in saving lives and protecting the human capital on which economies depend. Countries now face the dual challenge of controlling the current outbreak and rebuilding systems capable of developing human capital and meeting citizens' long-term expectations. As COVID-19 vaccines are being rolled out, governments can look toward a future after the pandemic and incorporate lessons from the crisis to build greater resilience into their systems. What steps should systems follow to build back better? What failings existed prior to the pandemic that can now be remedied? How can countries invest most strategically in services to protect and strengthen human capital for today and tomorrow?

Evidence to inform action

Tackling these challenges will require policy choices informed by evidence. Success in "building back better" will depend to an important degree on the quality and relevance of evidence available to inform policy makers' decisions and stakeholders' actions, including how to target resources for the highest returns. Therefore, measurement will play a crucial role in enabling countries to transform their health and education systems after the crisis.

Health and education are deeply intertwined in the process of human capital accumulation, and they are mutually reinforcing. For example, low-quality maternal, infant, and child health services put a child at risk for poor developmental outcomes, potentially leading to reduced learning and overall educational attainment, as well as worse health later in life (Currie 2009). Similarly, improvements in education, especially among women, have long been shown to boost health in families and communities. Reinforcing the evidence base on how schooling and health systems can work synergistically may accelerate human capital accumulation.

Learning from the Service Delivery Indicators surveys

There are many dimensions to health and education systems, including policies and guidelines, workforce recruitment and training, and incentives and pay.[1] Fundamental to system performance are the frontline settings where services meet citizens: at local schools, in clinics, and in hospitals. Comprehensive measurement of health and education services at the point of delivery can help to uncover bottlenecks to quality of care and education and, in turn, to human capital accumulation.

To advance this measurement agenda, a decade ago, the World Bank launched the Service Delivery Indicators (SDI) surveys. SDI surveys are nationally representative facility surveys that directly measure whether teachers know the material they are supposed to teach, whether health care providers are able to diagnose and treat common diseases, and whether schools and clinics have basic inputs like textbooks and stethoscopes (box 1.1). By documenting the competence and behavior of providers and the availability of inputs, SDI surveys offer a unique window into the quality of schooling and health care. Implemented systematically and with a core of comparable questions across countries and over time, SDI surveys allow cross-country benchmarking while speaking to the specificity of country contexts. The surveys are consistent in spirit with exercises like global indexes, which leverage competition across countries to trigger virtuous circles of debate and reform. Together with other international measurement initiatives, SDI surveys create a factual platform for dialogue around health and education reforms that engages a broad set of stakeholders, including governments, trade unions, parents, and patients.

As the SDI initiative continues to expand, this publication takes stock of more than a decade of SDI surveys in Africa. The SDI initiative began in 2008, when researchers and practitioners at the World Bank Group, in partnership with the African Economic Research Consortium and later supported by the

> **BOX 1.1** What do Service Delivery Indicators surveys measure?
>
> • *What do providers know?* Service Delivery Indicators (SDI) surveys directly measure teachers' knowledge by assessing how teachers answer questions on the fourth-grade material they are supposed to teach. Health workers are presented with patient case simulations ("vignettes") on symptoms of high-burden diseases such as malaria, diabetes, and pneumonia. Using the innovative methodology of vignettes, the SDI surveys are the only surveys of this scale that measure the knowledge of health care providers in a comparable way across countries.
>
> • *What do providers do?* Visiting unannounced, SDI survey enumerators assess the level of absence among teachers by observing whether classes are unattended and among health workers by cross-checking the presence of providers with the registry of personnel. With this pioneering approach, SDI surveys have generated the only internationally comparable data on provider absence in the education and health sectors, quantifying what many have only hypothesized.
>
> • *What do providers have to work with?* The quality of services is limited not only by the technical capacity and behavior of providers but also by physical resources—for example, access to clean water, electricity, and improved toilets. Although vaccines may be available in most health facilities, in many cases refrigerators cannot maintain the temperatures required to store those vaccines properly. In schools, enumerators visually check that children have textbooks with them and that each school has toilets separated by gender.
>
> • *What are children learning?* Establishing a direct link between accessing care in a clinic and patients' health status is complex. No single summary measure of health status is available in the literature, and, even if health status were easily and reliably measurable, after having seen a doctor, a patient could choose to be seen by a different doctor in a different clinic. Hence SDI surveys do not include measures of health outcomes. In contrast, the link between teachers' quality and knowledge and children's learning has been established in the literature. Moreover, children's numeracy and literacy can be measured in comparable ways. When fielded in schools, SDI surveys therefore include a measure of education outcomes—children's learning.

William and Flora Hewlett Foundation and the African Development Bank, worked together to develop novel survey tools and a methodology for measuring comprehensively the delivery of primary health care and primary school services. The first SDI health and education surveys were piloted in 2010 in Senegal and Tanzania. Since then, they have been scaled up to cover more than 15 African countries and adapted to settings outside Africa, such as Afghanistan, Armenia, Bhutan, Guatemala, Indonesia, Iraq, Moldova, and Pakistan, where work is ongoing or recently completed.

Are systems working for citizens?

SDI surveys adopt the perspective of an average patient or student, with a focus on indicators of provider presence, provider knowledge, and physical inputs that seek to proxy from different angles the quality of everyday services. The surveys are undertaken through enumerator visits to a representative sample of health facilities or schools in each country, including both public and

private facilities. SDI surveys report information about the availability and functionality of infrastructure, equipment, and other physical assets via enumerators' direct observation. In health, for example, survey enumerators ask about and observe the condition of water sources and sanitation facilities and whether health care providers have stethoscopes, antimalarial medicines, and other essential inputs to treat patients. During education surveys, enumerators check that blackboards are visible to all pupils and that students have a pen and paper to write with, as well as other materials for proper learning. Provider presence is measured in both health and education surveys with an unannounced visit to facilities, and providers' professional knowledge is directly tested. Finally, in schools, a sample of fourth-grade children is tested in math and language (both local vernacular and language of instruction).

SDI surveys are meant to complement routine information systems and provide greater depth. Routine information systems are the backbone of monitoring and evaluation in most countries and capture crucial information for the day-to-day management of systems, including tracking medicine stocks at health clinics and student attendance at schools. However, routine systems are unable to capture more detailed features, such as the knowledge and behavior of staff and multifaceted indicators of quality. In addition, self-reported routine data can be unreliable or missing altogether, as information management systems are still emerging in many of the countries that conduct SDI surveys.

Accountability for quality

The SDI initiative is premised on the concept of making services work for the poor, as outlined in the *World Development Report 2004* (World Bank 2004). This framing emphasizes the idea that measurement and information on providers' performance enable beneficiaries—students, parents, and patients—to demand better quality of services, thus holding frontline providers accountable. Using this "short route to accountability," health and education systems can be improved not just through government decisions but also through an active process of citizen engagement. This principle is echoed in a recent report of the Lancet Global Health Commission on High-Quality Health Systems, which notes, "Governments and civil society should ignite demand for quality in the population to empower people to hold systems accountable and actively seek high-quality care" (Kruk et al. 2018).

Low- and middle-income countries currently allocate roughly one-third of their budgets to human development sectors. Given the magnitude of this outlay, citizens should be engaged in demanding accountability for the efficient use of public resources and ensuring that these resources are distributed effectively and fairly. But, without consistent and accurate information on the quality of services, it is difficult for citizens or political leaders to assess how service providers are performing and to bring about improvements.

The relevance of the information provided by SDI surveys depends on how well what is being measured maps to the ultimate outcomes in a sector. *World Development Report 2018: Learning to Realize Education's Promise* offers a useful framework for thinking about how proximate inputs to education result in educational outcomes (World Bank 2018). On the health side, the Lancet Global Health Commission on High-Quality Health Systems developed a framework showing how processes, impacts, and inputs relate.[2] The Lancet Commission schema includes more detail than the World Development Report framework but can be mapped onto a similar arrangement of four elements: (1) the characteristics of service users (either students or patients); (2) physical inputs, such as equipment, supplies, and infrastructure in facilities; (3) the management of these facilities; and (4) the characteristics of providers, including their knowledge, capabilities, and behaviors. These relationships are depicted in figure 1.1. Although this simplified framework conveys well the "proximate causes" of health and education outcomes, in its concise version it does not make reference to systemic forces that shape these proximate factors, nor does it spell out the differential roles of key stakeholders, including policy makers, communities, the private sector, the legal system, and civil society organizations, among others. Nonetheless, these four main categories of the concise version align well

FIGURE 1.1   Inputs and outcomes in health and education
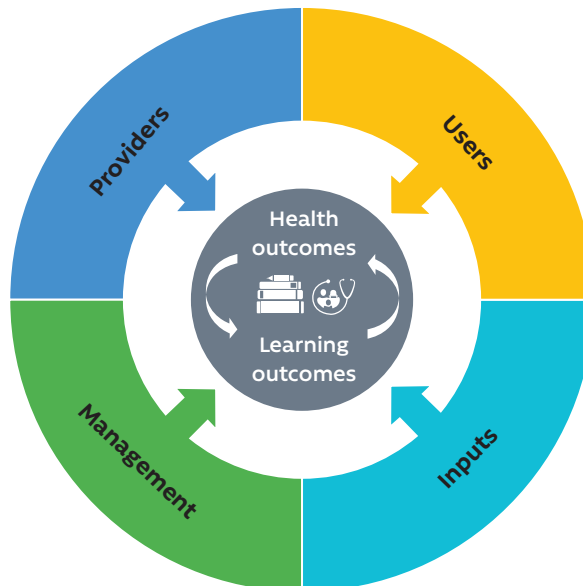
Source: Adapted from World Bank 2018.

with the SDI measurement approach and are useful for organizing the presentation of SDI analytical work. To date, SDI surveys have primarily measured physical inputs and provider characteristics and have been more limited in the measurement of user characteristics and outcomes. The SDI initiative has recognized a pending agenda to measure the quality of management and other key drivers of service quality (see chapter 4).

Aims and structure of the book

This book documents lessons learned from a decade of SDI surveys.[3] The surveys provide a nuanced view of the state of health care and education systems across Africa, showing both remarkable successes in certain aspects and the ongoing need for strengthening and revitalization in others. The book highlights the variation within and between countries, documenting how widely health care and education services may differ, depending on where they are sought, and offers elements for reflection on how health and education systems could be improved in a post-COVID-19 world.

Chapter 2 presents results from the SDI health surveys, with a focus on what the SDI surveys reveal about the current state of primary health care. It provides data on some of the most common obstacles and facilitators that patients face in seeking care, such as the availability of medicines or provider knowledge of common outpatient conditions. The discussion emphasizes the breadth of experience within countries and attempts to identify entry points to improve the provision of health care.

Chapter 3 discusses evidence from the SDI education surveys, which add a further dimension of measurement: learning outcomes, which can be analyzed in connection with teacher and school characteristics. The SDI education results highlight the substantial heterogeneity within and between schools and across countries, with interesting lessons to be derived from the comparison of public and private schools and the analysis of characteristics that distinguish the best- from the worst-performing schools.

The world of measurement is changing, and SDI surveys are changing with it. Measurement innovations in SDI surveys are discussed in chapter 4. Although the initiative began in Africa, SDI surveys have now expanded around the globe and continue to evolve in both content and form. Some changes have been guided by country-specific needs and evolving policy priorities, whereas other changes are premised on a need to understand the context in which health and education services are used. This richer contextual view expands the focus beyond the facility itself. This chapter details ongoing innovations and discusses the future goals of the SDI initiative.

Chapter 5 concludes, indicating how measurement can inform priorities for improving health and education systems. The COVID-19 crisis has tested systems in many countries to their limits, and this book is intended to guide thoughtful reforms as countries rebuild. The strength and resilience of health and education systems will be a major determinant of countries' success in the future, and the book makes a measurement-driven contribution to the policy discussions now being held.

Accompanying the analytical work of this book, the SDI team has curated and harmonized the historical SDI data collected since the initiative's inception. These data are accessible through its website, www.sdindicators.org, incorporating a newly designed interface that allows users to visualize interactively the SDI health and education indicators for all countries included in the book. The website, data repository, and data visualization tool will be updated periodically as more countries complete SDI surveys.

Notes

1 | The Systems Approach for Better Education Results (SABER) is a World Bank Group initiative that takes such a whole-of-system approach. SABER produces comparative data and knowledge on education policies and institutions, with the aim of helping countries to strengthen their education systems systematically and promoting learning for all.

2 | Kruk et al. (2018) propose a framework with three key domains: foundations, processes of care, and quality impacts. The foundations domain includes the following components: workforce, tools, governance, and the population. This is a similar framing to the World Development Report elements of providers, inputs, management, and users. The Lancet framework also includes three components of impacts on quality: better health, confidence in the system, and economic benefits, which are summarized as outcomes in the World Development Report framework.

3 | This book focuses on secondary analysis of SDI data collected over the last decade. It partially replicates some of the analysis undertaken in SDI country reports, albeit with slightly different assumptions in order to ensure consistency across countries.

References

Corral, P., and R. Gatti. 2020. "Accumulation Interrupted: COVID-19 and Human Capital among the Young." In *COVID-19 in Developing Economies*, edited by S. Djankov and U. Panizza, 286–302. London: Centre for Economic Policy Research.

Currie, J. 2009. "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." *Journal of Economic Literature* 47 (1): 87–122.

Elston, J. W. T., C. Cartwright, P. Ndumbi, and J. Wright. 2017. "The Health Impact of the 2014–15 Ebola Outbreak." *Public Health* 143 (February): 60–70.

Flabbi, L., and R. Gatti. 2018. "A Primer on Human Capital." Policy Research Working Paper 8309, World Bank, Washington, DC.

Human Rights Watch. 2020. "Impact of Covid-19 on Children's Education in Africa." Submission to the African Committee of Experts on the Rights and Welfare of the Child, 35th Ordinary Session, August 31, 2020.

Jain, R., A. Chopra, C. Falezan, M. Patel, and P. Dupas. 2020. "COVID-19 Related Immunization Disruptions in Rajasthan, India: A Retrospective Observational Study." Preprint manuscript. *medRxiv,* posted December 7, 2020. doi: https://doi.org/10.1101/2020.12.04.20244327.

Kim, J. 2018. "The Human Capital Gap: Getting Governments to Invest in People." *Foreign Affairs* (July-August).

Kruk, M. E., A. D. Gage, C. Arsenault, K. Jordan, H. H. Leslie, S. Roder-DeWan, O. Adeyi, et al. 2018. "High-Quality Health Systems in the Sustainable Development Goals Era: Time for a Revolution." *Lancet Global Health* 6 (11): e1196–e1252. doi: 10.1016/S2214-109X(18)30386-3.

Mahase, E. 2020. "Covid-19: What Do We Know about 'Long Covid'?" *British Medical Journal* 370: m2815.

Roberton, T., E. Carter, V. Chou, A. Stegmuller, B. Jackson, Y. Tam, T. Sawadogo-Lewis, and N. Walker. 2020. "Early Estimates of the Indirect Effects of the Coronavirus Pandemic on Maternal and Child Mortality in Low- and Middle-Income Countries." *Lancet Global Health* 8 (7): e901–e908. https://www.thelancet.com/action/showPdf?pii=S2214-109X%2820%2930229-1.

Sharma, J., K. Andrews, R. Conner, and R. Gatti. 2020. "What Can the Service Delivery Indicator Surveys Tell Us about COVID-19 Preparedness?" Policy Research Working Paper 9334, World Bank, Washington, DC.

UNICEF (United Nations Children's Fund). 2020. "UNICEF Calls for Averting a Lost Generation as COVID-19 Threatens to Cause Irreversible Harm to Children's Education, Nutrition and Well-Being." Press release, UNICEF, Paris, November 19, 2020. https://www.unicef.org/press-releases/unicef-calls-averting-lost-generation-covid-19-threatens-cause-irreversible-harm.

USAID (United States Agency for International Development). 2012. "Building Resilience to Recurrent Crisis." USAID Policy and Program Guidance, USAID, Washington, DC.

WHO (World Health Organization) 2020. "Pulse Survey on Continuity of Essential Health Services during the COVID-19 Pandemic: Interim Report, 27 August 2020." WHO, Geneva.

World Bank. 2004. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank.

World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank.

World Bank. 2020a. "The Human Capital Index 2020 Update: Human Capital in the Time of COVID-19." World Bank, Washington, DC.

World Bank. 2020b. "The COVID-19 Pandemic: Shocks to Education and Policy Responses." World Bank, Washington, DC.

World Bank. 2020c. *Poverty and Shared Prosperity Report 2020: Reversals of Fortune*. Washington, DC: World Bank.

2. Health service delivery in nine African countries

Background: An opportunity to transform primary health care

One of the most basic human aspirations is to live a long and healthy life. When health problems do arise, people want to receive good medical care close to their home. To meet these aspirations, governments and partners have worked to protect and promote health through primary health care (PHC) (UNICEF and WHO 1978). For four decades, PHC reform has been at the center of global efforts to expand access to health services and improve health outcomes. PHC values have been reflected in the Millennium Development Goals and Sustainable Development Goals (UN 2011; UN General Assembly 2015). Recently, country health leaders have reaffirmed their recognition that achieving universal health coverage depends on the ability of governments to expand PHC access to marginalized groups while ensuring high-quality care for all (WHO 2013a, 2013b; WHO, OECD, and World Bank 2018).

COVID-19 (coronavirus) has tested health systems in numerous ways, confirming what many in the health community have long known—strong PHC systems are the foundation of population health during and beyond health crises (WHO 2020). PHC systems are vital to control local outbreaks by preventing transmission, detecting cases, and ensuring quick clinical response. In addition, PHC systems need to ensure the continued delivery of nonpandemic essential health services to communities (PHCPI and World Bank 2020).

COVID-19 has also revealed weaknesses in many countries' PHC systems and aggravated existing inequities. For instance, preliminary evidence from 29 countries suggests that access to sexual and reproductive health services has decreased and violations of related rights have increased as an indirect result of COVID-19 (Endler et al. 2021). As in previous epidemics, substantial mortality may be attributable to the disruption of routine PHC services (Parpia et al. 2016 and Roberton et al. 2020). The strains on health systems have exacerbated underlying flaws, exposing weaknesses in supply chains and highlighting ongoing inequities in access to care. If the COVID-19 virus has created deep clefts in health care systems, it has often done so by widening and deepening the cracks and deficiencies already there.

As countries recover from the COVID shock and plan how to improve health system responses in the future, they will have a unique opportunity to strengthen and reform PHC, increasing investment while making it more fit for purpose. These reforms can be far-reaching and offer a chance to reimagine the health sector, strengthen links within it, and address neglected challenges. Measuring and ensuring quality remain vital for improving health systems, as estimates indicate that more deaths globally are due to low quality of health care (5.0 million deaths annually) than to lack of access to care (3.6 million deaths annually) (Kruk, Gage, Johnson et al. 2018). To achieve the greatest gains in population health, new resources will need to be invested wisely in health system reforms focused on increasing the quality and coverage of PHC. The Service Delivery Indicators (SDI) surveys offer insights for advancing health system reforms, with actionable indicators and a road map for improvement.

SDI health surveys: A finger on the pulse of primary health care

The SDI health surveys offer a set of indicators for benchmarking PHC performance. These indicators focus on potential determinants of the quality of PHC services: medical providers' level of knowledge; their efforts to provide patient care; and the availability of necessary equipment, supplies, and medicines. The distinctive strength of the SDI surveys comes from providing a "patient's eye view" of what does and does not work in the health system. To measure the various aspects of health system functioning, one can imagine a typical patient

progressing through the care-seeking process and the obstacles that she may face. For the patient to receive high-quality care, all parts of the system must be functional, and SDI surveys show the likelihood of shortfalls at each step of the process. Framing the analysis in this way emphasizes that the indicators of health system performance are not impersonal numbers. They mirror real people's experiences in the health system and quantify whether the patients who rely on the system receive the care they need. For that reason, these indicators provide crucial evidence that can be used to make the system more effective, responsive, and fair. This information is collected through enumerator visits to a sample of health facilities in each country.[1] The majority of the data collection is done through a survey administered to the facility manager as well as direct observation of the availability and functioning of infrastructure, equipment, medicines, and other physical assets.

Structure of this chapter

Consistent with the spirit of the SDI surveys, this chapter follows a typical patient in her care-seeking journey through the health system. It asks whether the patient will find health professionals present at her health facility and begins to explore whether the outpatient caseload may limit the ability of these health care providers to offer appropriate care. Then it summarizes evidence on whether health workers are trained and skilled to diagnose and treat the typical patient's condition correctly and whether they have the needed tools and medicines at hand.

The results presented suggest key opportunities for improving health care systems. Where provider absence is high, there is room for improving both monitoring and accountability, and, while some facilities have a large volume of patients, others are relatively underused, suggesting the possibility of improving care by shifting staff. Similarly, staff knowledge of basic conditions varies widely, and efforts could be made to strengthen competence among staff who form the frontline of care. Finally, some basic infrastructure, equipment, and medicines remain unavailable at many health facilities, constituting physical constraints to better care. In the wake of COVID-19, many health systems will undergo an overhaul, and this chapter concludes with recommendations for reimagining systems on the basis of findings across countries.

The findings presented in this chapter build on previously published work. A recent study using SDI health data assesses the quality of health care across Sub-Saharan Africa on the basis of providers' clinical attendance, knowledge of seven basic medical conditions, and availability of key medicines, reporting country-level averages for key indicators and an overall estimate of readiness to provide care (Di Giorgio et al. 2020). The study's findings emphasize the need

to strengthen the knowledge of health care providers, identifying this weakness as the major factor impeding further improvement in the provision of care. The results show how SDI tools and data can be used to diagnose key challenges in the health system. In related work, Andrews et al. (2021) examine the correlates of these key indicators to assess differences in the performance of health facilities within and between countries. Their paper analyzes levels and trends in the provision of primary care services and offers a comprehensive assessment of both the heterogeneity of care and the systems-level factors that may be driving inequities.

The analysis presented in this chapter extends these investigations, presenting results parallel to those in chapter 3 on education and allowing for synthesis of findings across sectors.

Sample, methods, and framework

The SDI health surveys have been implemented for more than 10 years across 13 countries in Sub-Saharan Africa, and this chapter presents data from across this period.[2] In countries with multiple SDI surveys, only the most recent survey is included, because data from the same country over time are not fully comparable due to differences in sampling and measurement strategies.[3] The resulting data set encompasses information on nine countries, covering 7,810 health facilities. It includes results from the following country surveys: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria[4] (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013).

In each country, the sample of surveyed facilities is drawn from the ministry of health list of all health facilities offering primary care services. This list includes public facilities and facilities operated and administered by private entities or nongovernmental organizations (NGOs). Given the surveys' focus on primary care, facilities at all levels of care, such as hospitals, health clinics, and health posts (or the national equivalent), are included if they provide primary care services. Sample selection is stratified by urban-rural location and by type of facility. The proportion of health posts varies, from none in Mozambique to 74 percent of all facilities in Sierra Leone, and the proportion of rural facilities also varies, from 49 percent of the sample in Madagascar to 88 percent in Mozambique. All surveys were designed to be nationally representative, except for the survey in Nigeria, where data were collected in 12 of 36 states because of logistical constraints. In Kenya, data are representative not only at the national level but also at the county level.[5] Details of the sample are presented in table 2.1.

The SDI surveys purposefully include facilities operated by nongovernmental entities, in line with the goal of providing an inclusive assessment of citizens' access to care. Because the private sector is responsible for a substantial portion

TABLE 2.1 Size of the SDI health sample in nine African countries, by country

Country	Survey year	Health facilities	Health care providers	Clinical vignette interviews
Kenya	2018	3,038	24,404	4,485
Madagascar	2016	444	2,200	619
Mozambique	2014	195	2,972	694
Niger	2015	255	1,331	594
Nigeria	2013	2,385	21,318	5,017
Sierra Leone	2018	536	5,055	829
Tanzania	2016	383	5,160	498
Togo	2013	180	1,364	527
Uganda	2013	394	2,347	733
All		7,810	66,151	13,996

Source: SDI (Service Delivery Indicators) health surveys.
Note: This table shows the number of health facilities, health care providers on the roster, and health care providers given a clinical vignette assessment, by country. A subsample of up to 10 providers at each facility was selected for the clinical vignette interview, taken from the providers who report regularly providing outpatient care.

of health care delivery in many countries, it is important to understand the current state of the private system and any unique challenges that it may face. In the SDI sample, the private system includes private for-profit facilities (15.9 percent of sample), faith-based nonprofit facilities (6.5 percent), NGO facilities (2.0 percent), and community-run organizations (1.3 percent). The overall share of private facilities varies, from 1 percent of the sample in Mozambique to 42 percent of the sample in Kenya. Summary statistics and estimates are reported using facility-level survey weights, which are calculated on the basis of the inverse probability of being sampled.[6] In all figures, countries are ordered by increasing gross domestic product (GDP) per capita to visualize how performance on health service indicators may correlate with country income.[7]

The indicators operationalized in this retrospective analysis are based on the understanding that "quality of care is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge" (WHO, OECD, and World Bank 2018). Although the surveys do not comprehensively measure all aspects of quality encompassed by this definition, the data obtained cover eight key indicators: health care provider absence rate, outpatient caseload per health care provider, diagnostic accuracy, treatment accuracy, management of maternal and neonatal complications, drug availability, equipment availability, and infrastructure availability. To underscore what these indicators mean for the patient's experience, this chapter frames them as a series of questions.

Explicitly or implicitly, the following questions are on a patient's mind as she engages with the health system:

- *Provider's effort*. Will health care providers be present at the health facility? Will health care providers be too busy with other patients?

- *Provider's knowledge*. Are health care providers able to diagnose and treat common conditions correctly?

- *Inputs*. Will the necessary infrastructure, equipment, supplies, and medicines be available?

These questions and the indicators embedded in them provide a broad overview of the current state of the health system, with each indicator measuring a different aspect of quality of care.[8]

Will health care providers be present in the health facility?

After deciding to seek care at a health facility of her choice and overcoming any barriers to reaching the facility, a patient would expect health care professionals to be available at the health facility during scheduled work hours. However, even among workers who are assigned to be on duty, it is not guaranteed that every health care professional will be present and available to provide services to patients. Health worker absence has long been recognized as a fundamental obstacle to improving the quality of care in low- and middle-income settings (Bamgboye and Adeleye 1992; Belita, Mbindyo, and English 2013; Chaudhury and Hammer 2004). Despite recent gains, the problem persists in many countries, and both a better understanding of the underlying drivers and innovative solutions are needed to address it (Tumlinson et al. 2019). The SDI surveys reveal that, on average, 6 out of 10 health care providers are present at the facilities surveyed on any given day.

Across all nine countries, 43 percent of providers are absent from their facility during an unannounced visit. This high absence rate includes absence for both authorized and unauthorized reasons (further analyzed below) and is consistent with rates observed in Bangladesh, Ecuador, India, Indonesia, Peru, and Uganda, where 35 percent of health workers, on average, are absent from health clinics during an unannounced visit (Chaudhury et al. 2006). SDI surveys include two visits, of which the second is not announced, to assess staff presence, allowing for an unbiased estimate of the absence rate on a typical day of operation. Absence rates differ substantially across countries, ranging from more than 50 percent in public facilities in Kenya to less than 20 percent in private facilities in Mozambique and Tanzania, as shown in figure 2.1. Outside of

FIGURE 2.1   Absence rate of care providers in nine African countries, by country and health facility ownership
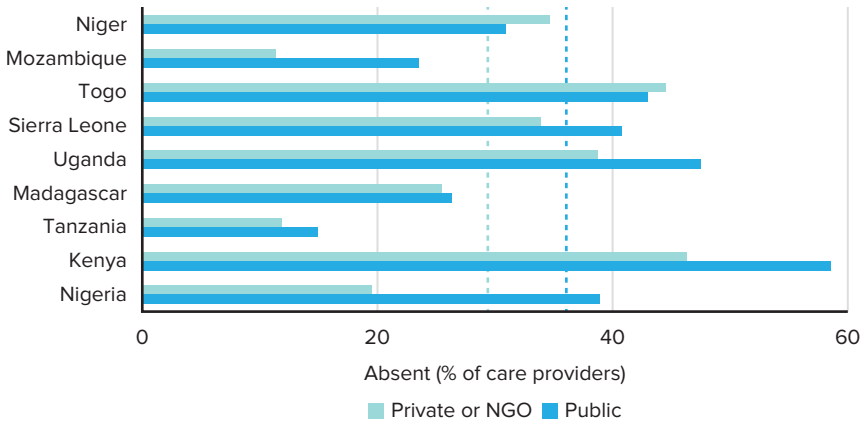
Source: SDI (Service Delivery Indicators) health surveys.
Note: This figure shows the total rate of absence during a second, unannounced visit. The denominator is up to 10 randomly selected health care providers listed on the roster at each facility. Dashed lines are overall average. Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013). NGO = nongovernmental organization.

Niger and Togo, more health care providers are absent in public facilities than in private ones.

Health care providers may be absent from the facility for a variety of reasons. Figure 2.2 shows the reasons for absence in public facilities, including training or a meeting (21 percent), official mission (11 percent of absences), sick or maternity leave (11 percent), and other authorized absence (28 percent). Across all public facilities, only 4 percent of providers are absent without authorization, with the highest rates among doctors and clinical officers (5 percent) and nurses (4 percent) and the lowest rates among other health workers (3 percent).[9] Overall, more than 90 percent of absences in public facilities are authorized. However, this is a likely overestimate because facility administrators may be hesitant to report unauthorized absences, and there is no method for confirming the reasons for absence.

Notwithstanding important contextual differences, these data indicate some general trends across countries. The total absence rate is high, exceeding 20 percent in almost all countries. Unauthorized absences are a much smaller fraction, but they are more prevalent in public than in private facilities in all nine countries. The higher rate of absence in public facilities may stem from differences in incentive structures, differences in management and supervision, or differences in demands that take public health care providers outside of the facility. Some absences may be out of the control of facility administrators,

FIGURE 2.2 Reasons for absence of health care providers in public health facilities in nine African countries, by country
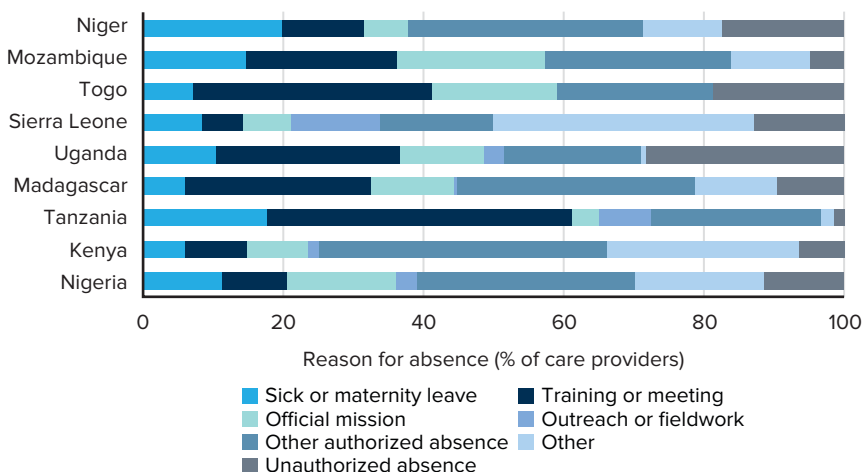
Source: SDI (Service Delivery Indicators) health surveys.
Note: This figure shows the reported reasons for absence among health care providers listed as absent in public facilities. The reasons for absence differ by country and are grouped into major categories. Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013).

and national-level strategies could be considered to keep staff on-site, such as limiting the frequency of government-mandated trainings or discouraging staff from maintaining secondary employment.

Will health care providers be too busy with other patients?

For primary care, when patients come to a health facility, they hope that a medical professional will be available and that they will not have to wait inordinately long to receive care. Similarly, health care providers may hope that they will be able to see all of the patients who present in a day and that they will not be overburdened with more patients than they can reasonably triage. To estimate the burden of care on staff, SDI surveys measure caseload, defined as the number of outpatient visits per health care provider per day.

The use of this metric in the broader health literature arises from recognition of the uneven distribution of health workers and the mismatch between the distribution of workers and health care needs. Countries with the lowest relative needs often have the highest number of health workers, whereas countries with the greatest burden of disease have a much smaller health workforce. Specifically,

the World Health Organization (WHO) estimates that Sub-Saharan Africa bears more than 24 percent of the global burden of disease but employs only 3 percent of health workers (WHO 2006). From a financial perspective, staff costs, in the form of salaries, allowances, and benefits, often account for more than half of total health system costs (Kaplan et al. 2014). Getting staff in the right places is therefore crucial for ensuring a well-run and cost-effective system.

The SDI surveys indicate that, on average, a health care provider in these countries attends roughly 13 outpatients per day. Caseloads range from a low of 3 outpatients per provider per day in Nigeria to a high of 23 in Kenya and Mozambique. On average, providers in public facilities attend about 14 patients per day, whereas providers in private facilities attend about 12 patients per day. The pattern varies by rural-urban location, depending on the country, as shown in figure 2.3.

Although there is no "ideal" number of patients per provider per day (Speakman 2016), the estimates from the SDI surveys suggest that outpatient caseload is low at many facilities in Sub-Saharan Africa. The share of facilities with low caseloads (defined here as fewer than 5 outpatients per provider per day) ranges from 3 percent of public facilities in Uganda to 87 percent in Nigeria (figure 2.4). Some of these differences may be due to differences in the sample composition between

FIGURE 2.3 Caseloads in public health facilities in nine African countries, by country and urban-rural location

Patients per provider per day

■ Urban ■ Rural
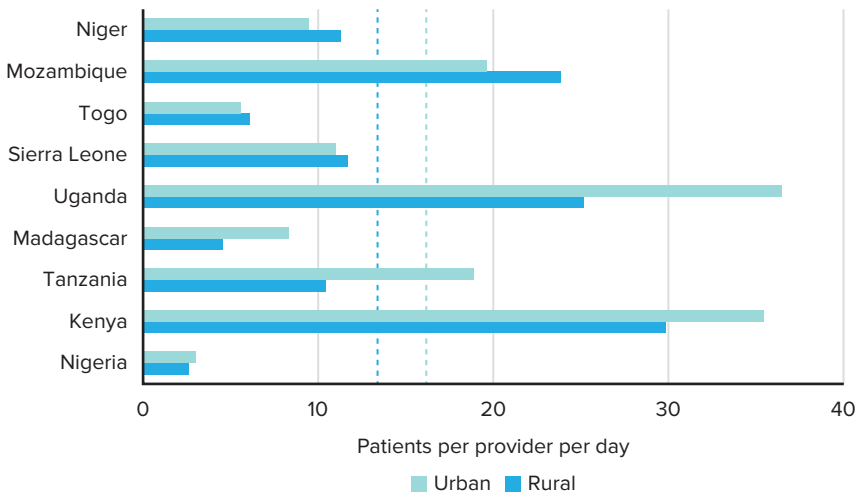
Source: SDI (Service Delivery Indicators) health surveys.
Note: Caseload is based on the estimated number of outpatients per day (measured as the total number of outpatients in the prior three months), divided by the number of medical staff at the facility on a typical day. Dashed lines are overall averages. Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013).

FIGURE 2.4 Share of public health facilities with fewer than five patient visits per provider per day in nine African countries, by country
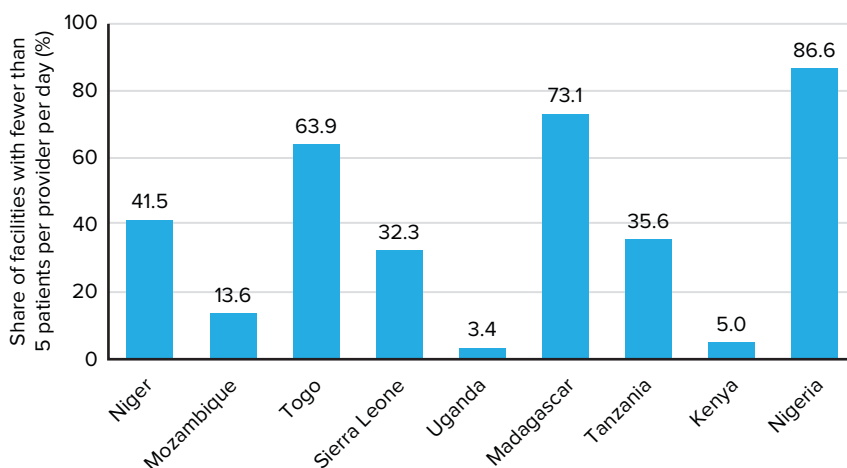
Source: SDI (Service Delivery Indicators) health surveys.
Note: The cutoff of five visits per provider per day is intended to represent a quantity of visits that can be handled easily in a routine workday. Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013).

countries, with Nigeria including many lower-level facilities that appear to be visited less frequently by patients. However, among public facilities overall, almost half (45 percent) have an outpatient caseload below 5 patients per provider per day. In contrast, in high-income economies like the United States, an average physician attends approximately 20.2 patients per day (Hawkins 2012).

The estimates of patients per provider per day are corroborated by health management information system (HMIS) data from Kenya. Extracting data from Kenya's District Health Information Software (DHIS2) for the same time period as the SDI surveys reveals a similar overall volume of outpatients as in the SDI survey; further details are shown in box 2.1. These findings are in line with recent analyses in Latin America, which look at efficient allocation of care. In Ecuador, a World Bank report estimates that approximately one-third of primary care facilities have low case counts (fewer than 500 outpatient visits per month) (Vermeersch and Giovagnoli 2020). Similarly, another analysis finds that about 10 percent of health facilities in Peru have fewer than 30 patient visits per day (World Bank 2020b). Although it is difficult to determine the optimal theoretical caseload, in both of these countries other similarly equipped facilities see three to four times the volume of patients.

Caseloads reflect only the direct patient interaction aspects of the health care providers' job; they do not reflect responsibilities such as administrative and management tasks (maintaining medical records, managing procurement

---

BOX 2.1   Triangulating SDI survey findings with administrative data

Caseload estimates in the Service Delivery Indicators (SDI) surveys are slightly lower than expected, at odds with researchers' initial hypothesis and the literature on shortages of human resources in health care (Liu et al. 2016). As such, it is important to confirm these findings using an alternative source of data. Many countries routinely collect information on the number of patients seen at each facility through a health management information system (HMIS). The collection of monthly outpatient values in the Kenya District Health Information Software (DHIS2) allows comparison with SDI survey estimates. DHIS2 has some limitations, such as an absence of information on staff counts and facility hours, and therefore the

comparison is based only on the reported number of outpatients. These data were extracted from DHIS2 for the same period as the SDI survey (January to March 2018), and SDI-surveyed facilities were matched to DHIS2 data using facility name and district. Data were missing or zero in DHIS2 for 439 of 3,034 facilities (14.5 percent).

Figure B2.1.1 depicts a simple scatter plot of the sum of outpatients as recorded by DHIS2 versus the SDI estimates. The overall correlation is 0.79. Although the caseload findings in the SDI surveys are approximately corroborated with data from routine systems, there is substantial variation in the quantity of patients seen, even at relatively similar facilities.

FIGURE B2.1.1   Number of outpatients in DHIS2 versus SDI surveys, with line of equivalency and best fit, Kenya, 2018
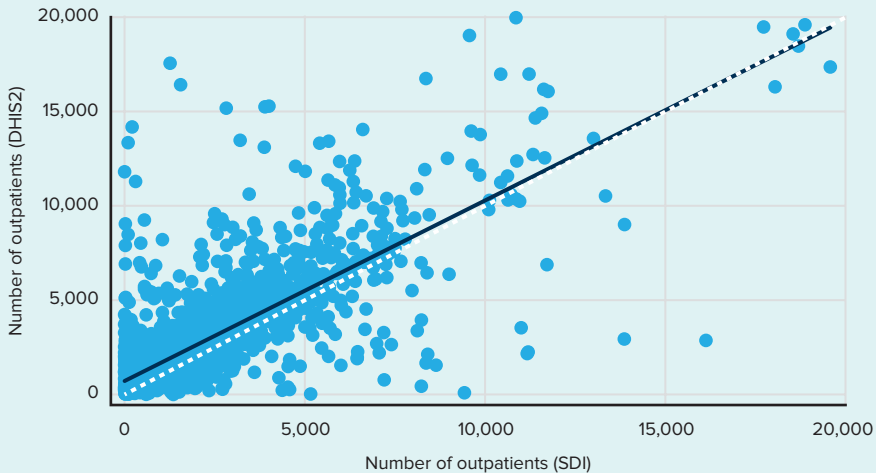
Source: DHIS2 (District Health Information Software) data; SDI (Service Delivery Indicators) health surveys.
Note: Each dot represents one facility, with the DHIS2-reported outpatients on the y-axis and the SDI-reported outpatients on the x-axis. The solid black line is a line of equivalency (x = y), and the dashed white line is the line of best fit.

---

systems and supply stocks, training or managing other health care providers). In addition, the caseload estimates are based on the number of outpatient visits seen in the facility in the prior three months and the total number of medical staff working at the facility, so they may not adequately capture the variation in day-to-day operations. The low caseloads observed therefore do not necessarily suggest

that health care providers have ample time to see more patients, because their time may be stretched to accommodate other activities, including providing care to inpatients (though the health facilities included in the SDI samples are mainly primary care facilities). However, they do suggest that some commonly reported problems, such as overcrowding and long wait times, may be due to factors such as poor facility management, large administrative burdens, or uneven distribution of patients throughout the day rather than a lack of sufficient staff.

Will health care providers be ready to provide quality care?

A patient's basic expectation is that health care providers will exhibit clinical competence in providing care. But competent care, where the health care provider accurately diagnoses and appropriately treats illnesses, is not a given. SDI data can shed light on facets of the patient experience that have to do with the clinical skills of health care providers. Will the providers competently assess the patient's condition, ask relevant questions, perform appropriate tests, and recommend suitable treatment?

In recent years, quality of care has received more attention in the health research community, with increasing recognition that good health outcomes depend not just on patients' access to care but also on the competence and skill of the health care provider. Accurate diagnosis and treatment are important for the health outcomes of patients and can also influence future patterns of health care use (Escamilla et al. 2018; Rao and Sheffel 2018). The SDI survey includes clinical vignettes that are administered to health care providers. This innovative addition measures the quality of clinical care, unlike the inputs-focused perspective taken in many earlier surveys (Das and Leonard 2006). Clinical vignettes may be less reliable for assessing quality of care than other methods, such as the use of standardized patients, but they are easier to implement, less expensive, and less disruptive to health facility operations. Overall, clinical vignettes have been shown to be "a valid and comprehensive method that directly focuses on the process of care provided in actual clinical practice" (Peabody et al. 2000).

In the SDI surveys considered in this book, health care providers are tested on five core vignettes: childhood diarrhea with dehydration, childhood pneumonia, adult tuberculosis, adult diabetes mellitus, and childhood malaria with anemia. Additionally, countries may add specific vignettes and occasionally remove vignettes (for example, the malaria with anemia vignette was not administered in Kenya). These vignettes represent common clinical cases that a health care provider would face in the low- and middle-income-country context.

These high-burden conditions make up 30 percent of all-age disability-adjusted life years in Sub-Saharan Africa.[10] Each provider is scored on the percentage of vignettes for which he or she provides the correct diagnosis and treatment.[11] Multivariate regressions to test the relationship between provider-level variables are described here and presented fully in a previous paper.[12] Further details on the vignettes are available in appendix A, tables A.5 and A.6.

Clinical vignettes are useful for measuring the diagnostic and treatment accuracy of health care providers. They also provide valuable information on adherence to clinical protocols. The vignettes contain country-adapted information to simulate a full consultation, including recommended questions on a patient's history, physical examination, laboratory tests, and options for care, which allows for a full measurement of provider adherence to clinical guidelines. For example, in Niger, the results of the clinical vignettes show that only 3.5 percent of health care providers accurately diagnose diarrhea with severe dehydration. To assess the severity of the case, the WHO guidelines for the integrated management of childhood illness (IMCI) recommend administering a skinfold test, checking for lethargy, checking whether the child is able to drink, and checking for agitated or irritable behavior. IMCI requires two positive danger signs to denote a case as severe, but only 24 percent of providers in the sample checked for both danger signs. Health care providers in Niger who inquire about each of these signs are significantly more likely to arrive at a correct diagnosis. The vignettes offer rich data, showing gaps in knowledge that can improve the understanding of provider performance and offer insights for policy or investment actions.

Figure 2.5 shows the estimated confidence intervals for diagnostic accuracy by type of provider and country. Overall mean diagnostic accuracy (the percentage of all vignettes administered for which health care providers give correct diagnoses) varies by country, from a high of 69 percent in Tanzania to a low of 40 percent in Nigeria. Across the sample, doctors and clinical officers have the best diagnostic accuracy (67 percent), followed by nurses (55 percent) and other medical staff (36 percent). However, the range by type of provider varies substantially by country, from a 10-percentage-point difference between doctors and other medical staff in Mozambique to a 43-percentage-point difference in Uganda. Controlling for both facility- and provider-level characteristics, doctors have the highest diagnostic accuracy, males have higher diagnostic accuracy than females, and adults ages 40–49 years have slightly higher diagnostic accuracy than younger or older colleagues. Perhaps not unexpectedly, providers with secondary and postsecondary education perform significantly better than those with only primary education. There is no significant difference between the performance of providers in urban and rural facilities or between those in public or private facilities, but providers at health posts score significantly lower than those at hospitals or health clinics. Because less specialized providers will likely continue to provide the bulk of diagnoses at

FIGURE 2.5 Diagnostic accuracy in nine African countries, by country and type of health care provider (95% confidence intervals)

Accuracy (% of correct diagnoses)

■ Doctor ■ Nurse ■ Other
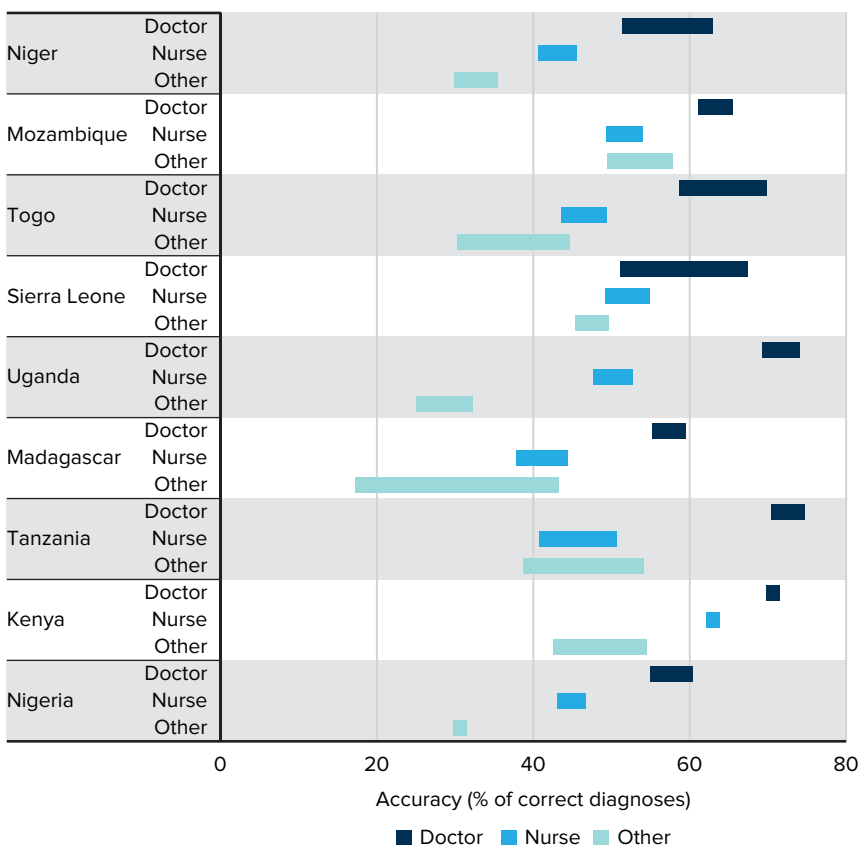
Source: SDI (Service Delivery Indicators) health surveys.
Note: Diagnostic accuracy is combined across the five core vignettes, shown in figure 2.6. The ranges shown represent 95% confidence intervals. Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013).

frontline facilities, these results suggest that countries could do more to build competencies and support the ongoing training of these workers.

Diagnostic and treatment accuracy results are shown in figure 2.6. The accuracy of health care providers in diagnosis varies among the five disease vignettes: tuberculosis (86 percent correct), diabetes (60 percent), pneumonia (59 percent), diarrhea with dehydration (27 percent), and malaria with anemia (22 percent). Treatment is correct in more than 50 percent of cases for all diseases, except malaria with anemia, which has a correct treatment rate of 22.7 percent. Doctors and clinical officers are more likely to perform better in treatment accuracy than nurses and other staff, providers with postsecondary education perform better, and male providers perform

FIGURE 2.6 Health care provider accuracy of diagnosis and treatment in public health facilities in nine African countries, by disease
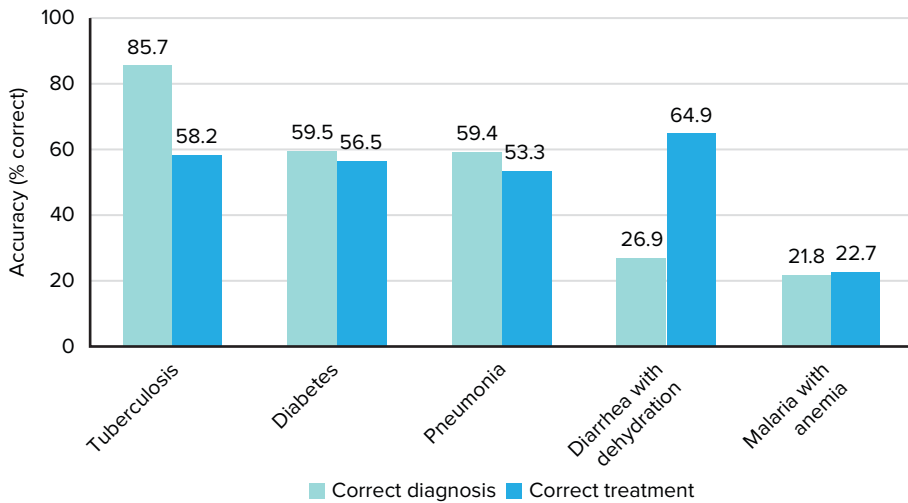
Source: SDI (Service Delivery Indicators) health surveys.
Note: Correct treatment is possible even without correct diagnosis and therefore can be higher, as in the case of diarrhea with dehydration and malaria with anemia. Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013).

better overall. Treatment accuracy is higher at hospitals than at other types of facilities, higher at rural than at urban facilities, and higher at public than at private facilities. Treatment accuracy is higher than diagnosis accuracy for the two dual-diagnosis conditions because providers sometimes fail to diagnose comorbidity (for example, diagnosing malaria but not anemia) but offer treatment that is satisfactory for both conditions (for example, prescribing artemisinin combination therapy and iron supplements).[13] These results suggest that patients are likely to receive insufficient treatment, particularly where multiple conditions such as diarrhea with dehydration are present, and that this problem is more common when treatment is provided by less specialized staff.

To offer information about the management of urgent maternal and child health conditions, SDI surveys include two additional vignettes, one on neonatal asphyxia and one on postpartum hemorrhage. Both conditions have a clear set of guidelines and recommended actions that providers should undertake to reduce the risk of maternal and neonatal morbidity and mortality. For the purposes of scoring across countries, providers are assessed on seven actions for neonatal asphyxia (call for help, place the baby in a neutral position, check the baby's heart rate, check the baby's breathing, dry the baby, keep the baby warm, and initiate resuscitation with a bag or mask) and five actions for postpartum hemorrhage (determine the cause, provide bimanual uterine massage, place a

foley catheter, run an intravenous line, and provide oxytocin or similar drugs). Providers are scored on the number of these correct actions that they propose. Results for neonatal asphyxia by type of provider are shown in figure 2.7. For these two measures, nurses score as well as doctors or clinical officers on the neonatal asphyxia vignette and almost as well on the postpartum hemorrhage vignette. Other medical staff, older age groups, and males score worse on both measures. These results suggest that the average provider can identify only half of the necessary actions in an emergency situation, although it is encouraging that the scores of nurses are comparable to those of doctors on this measure.

FIGURE 2.7 Actions taken for neonatal asphyxia in public health facilities in seven African countries, by country and type of health care provider (95% confidence intervals)
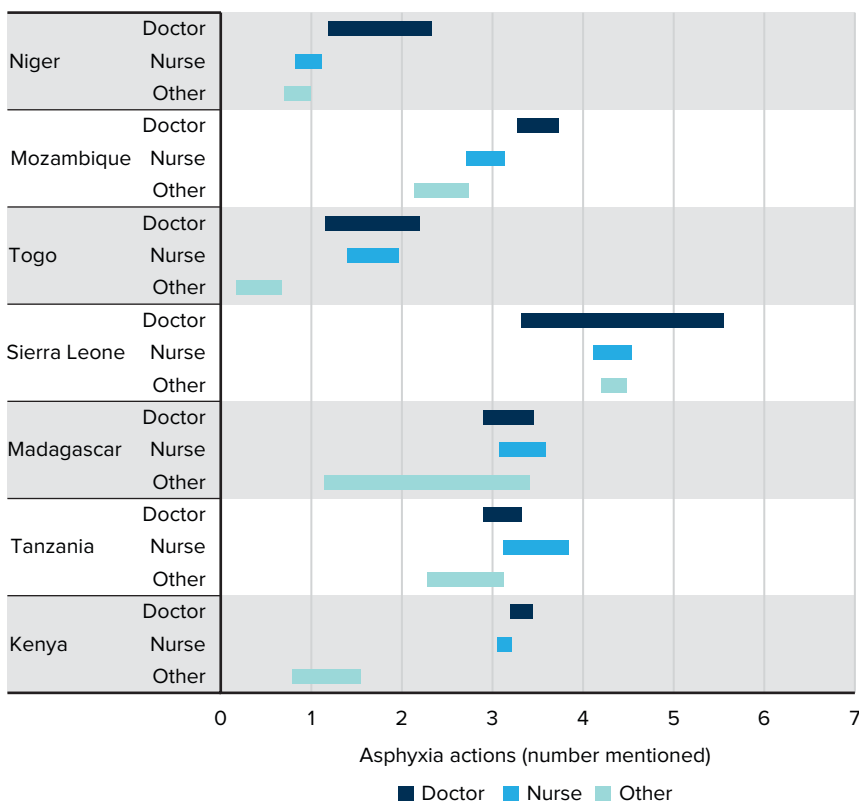
Source: SDI (Service Delivery Indicators) health surveys.
Note: Health care providers are graded on the basis of the number of actions that they mention (out of seven possible actions) for responding to a case of neonatal asphyxia. The ranges shown represent 95% confidence intervals (black, blue, and green bars) for the average number of actions recommended. Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Sierra Leone (2018), Tanzania (2016), and Togo (2013). Nigeria and Uganda are omitted because of incomparability of the standards used in assessment.

The SDI survey also collects information on inappropriate use of antibiotics, an ongoing global challenge that can contribute to the rise of antibiotic resistance (Laxminarayan et al. 2013). In the SDI surveys, inappropriate antibiotic prescription is defined as prescribing an antibiotic for the tuberculosis vignette (aside from the antibiotics recommended as part of the tuberculosis regimen) or any antibiotics for the diarrhea vignette (for which antibiotics are not indicated by examining the patient). Inappropriate antibiotic use is calculated as the percentage of health care providers who inappropriately prescribe antibiotics among all health care providers in the clinical vignettes. Results in figure 2.8, indicate that nearly half of health care providers prescribe an antibiotic in cases

FIGURE 2.8 Inappropriate antibiotic recommendation in public health facilities in nine African countries, by country and type of health care provider (95% confidence intervals)
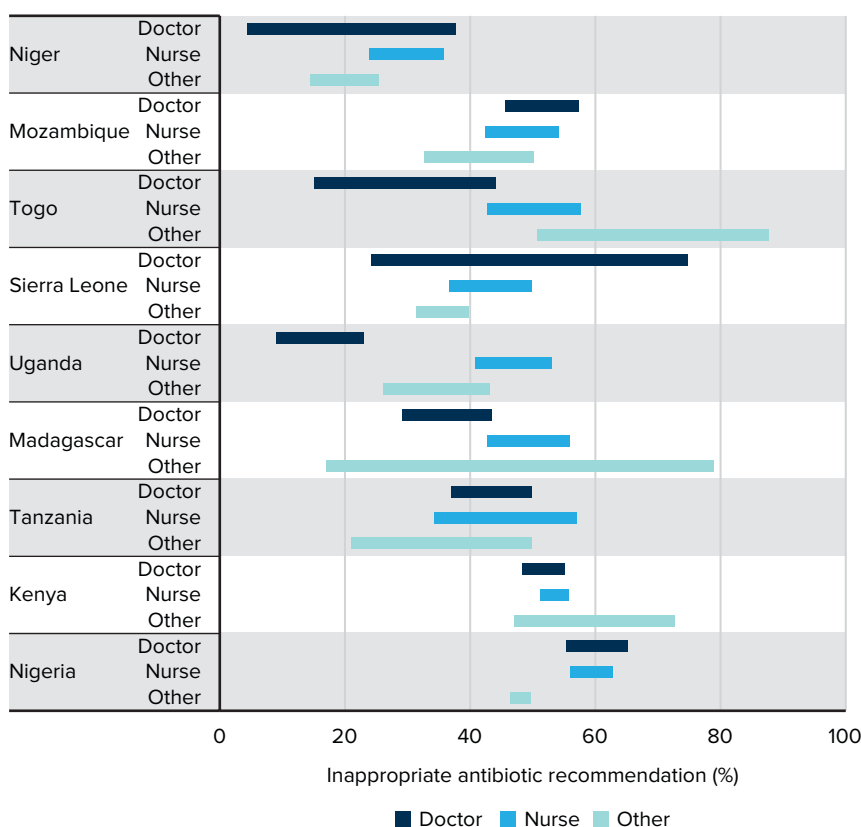
Source: SDI (Service Delivery Indicators) health surveys.
Note: Inappropriate antibiotic prescription is defined as prescribing an antibiotic for the tuberculosis vignette (aside from the antibiotics recommended as part of the tuberculosis regimen) or any antibiotics for the diarrhea vignette (for which antibiotics are not indicated by examining the patient). The ranges shown represent 95% confidence intervals (black, blue, and green bars) for the % of health care providers giving an inappropriate antibiotic prescription. Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013).

where it is not recommended. There is variation by type of provider in prescription patterns, with doctors more likely to prescribe antibiotics inappropriately in Mozambique, Nigeria, and Sierra Leone, whereas nurses or other medical staff are more likely to prescribe antibiotics inappropriately in other countries. Inappropriate prescription of antibiotics is not noticeably different between providers at public versus private facilities or at rural versus urban facilities, but it is higher at hospitals than at health centers or health posts. Health care providers over the age of 50 are less likely than younger providers to prescribe antibiotics inappropriately, and males are more likely to prescribe antibiotics inappropriately. In some cases, antibiotics are ordered inappropriately in addition to correct treatment, which still suggests widespread overprescription of antibiotics. The variation suggests that health care providers with more education or training are not always more likely to make better decisions regarding prescriptions. Although clear guidelines and oversight of antibiotic use could help to reduce overprescription, the recent literature identifies knowledge gaps, misaligned incentives, and patients' own demand as possible causes of overprescription (Lopez, Sautmann, and Schaner 2021).

Will the necessary infrastructure, equipment, supplies, and medicines be available?

Even when a patient is treated by health care providers who have high clinical competence, the availability of key inputs such as basic infrastructure, equipment, clean water, supplies, and medicines may constrain the quality of care. Lack of clean water, sanitation, and hygiene facilities has been well documented as a risk for increased infection in health care settings; but global estimates suggest that these deficiencies remain the norm at up to a quarter of health facilities (UNICEF 2019). The availability of these physical resources does not guarantee system competence, but any gap in the provision of these structural inputs to care is a limiting factor in optimizing service delivery, regardless of the technical ability and behavior of health care providers (Leslie, Sun, and Kruk 2017).

Infrastructure availability

Basic infrastructure availability—defined as the availability of an improved water source, improved toilet, and electricity[14]—varies from a country average of 77 percent in Kenya to 21 percent in Niger. Infrastructure availability is significantly higher in urban areas, driven partly by the higher rates of electrification. Infrastructure availability is also significantly higher at private facilities than at public facilities across all countries and, predictably, is highest at hospitals and lowest at health posts.

Figure 2.9 shows infrastructure availability by country and the urban-rural differentials. In all countries, urban facilities have much greater infrastructure availability, but this gap varies across the sample. Niger has the greatest gap, with 85 percent of urban facilities having basic infrastructure, compared to 18 percent of rural facilities. The gap between urban and rural facilities is wider in countries with lower national income, like Madagascar and Sierra Leone, than in richer countries, like Kenya and Tanzania. Figure 2.10 highlights which components are lacking in facilities and shows when multiple components are missing. To understand and design policies to address low infrastructure availability, it is useful to decompose the gap in each country. For instance, in Kenya, relatively few facilities lack more than one item of basic infrastructure, whereas in Niger a greater proportion of facilities have multiple deficits. Similarly, the SDI results reveal that lack of access to electricity is a problem in all countries, but it is most pronounced in Uganda, whereas lack of improved toilet and improved water (each separately) are more common in Mozambique and Tanzania, respectively. Figure 2.10 also provides actionable evidence. For example, the results suggest that addressing only one missing piece of infrastructure in facilities in Uganda could increase the fraction of facilities with all three items from under 60 percent to almost 95 percent.

Given the variation within and between countries, what are the strongest predictors of basic infrastructure availability? Multivariate regressions show

FIGURE 2.9 Availability of basic infrastructure in public health facilities in nine African countries, by country and urban-rural location
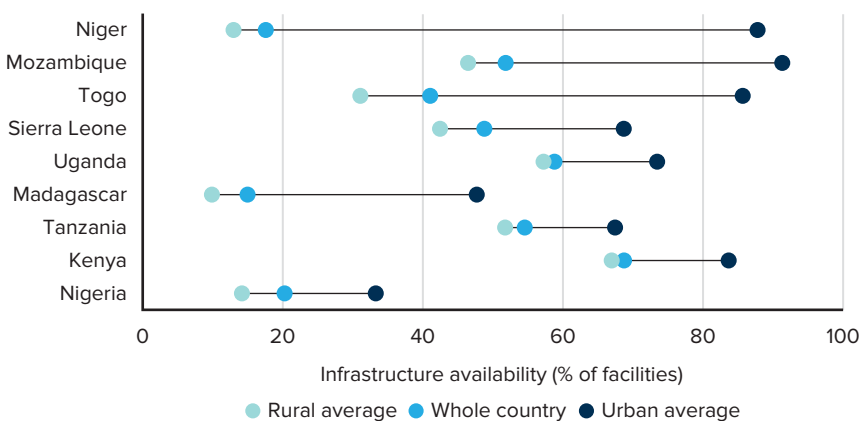
Source: SDI (Service Delivery Indicators) health surveys.
Note: This figure shows the percentage of facilities with all basic infrastructure available (improved water, improved sanitation, and electricity). Light green dots represent the average score at rural facilities, black dots represent the average score at urban facilities, and blue dots indicate the whole-country estimate. Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013).

FIGURE 2.10   Availability of basic infrastructure in public health facilities in nine African countries, by country and type of infrastructure
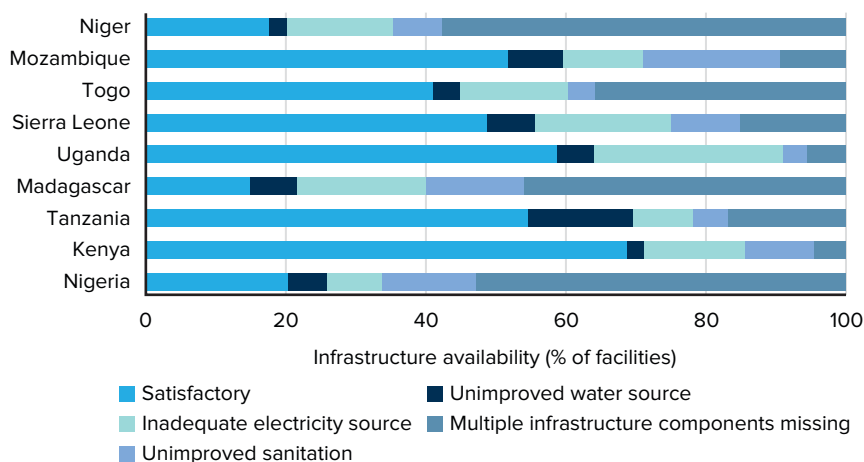
Source: SDI (Service Delivery Indicators) health surveys.
Note: This figure shows the breakdown of infrastructure (improved water, improved sanitation, electricity) not available in public facilities by country. "Multiple missing" refers to facilities that have more than one piece of infrastructure unavailable. Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013).

that country, level of health facility, facility ownership, and urban location are all important determinants of basic infrastructure availability. The urban-rural gap is particularly notable and provides important suggestive evidence that rural populations, which are typically reliant on public health care, have an inequitable share of facilities without access to improved water, sanitation, and electricity. In Kenya, infrastructure availability is lower in counties with a lower household wealth index, as shown in box 2.2. Paying attention to within-country inequities therefore must be a priority, as countries devise plans to overcome insufficiencies in the foundations needed to provide basic clinical care.

Equipment and medicine availability

A patient coming to a health facility would also hope that the facility is stocked with all of the tools and medicines needed for proper diagnosis and treatment. The list of necessary equipment and medicines can vary depending on the level of facility and type of care provided, but the SDI survey focuses on basic items. For equipment, basic tools are items such as a thermometer, stethoscope, blood pressure cuff, and weighing scale. These items represent the bare minimum that should be present in all facilities offering preventive and curative health care services. For medicines, the survey focuses on 14 common medicines,

BOX 2.2   Facility characteristics and wealth: Evidence from Kenya

Policy makers have pursued equity in health care provision for many decades, acknowledging that health systems should aim not just to maximize coverage but also to ensure that care is available for those who need it most (WHO 2008). By measuring the variation in health facility characteristics within a country, the SDI surveys can show the relative equality or inequality of care provision. An analysis undertaken with data from Kenya demonstrates this process. Matching the 2018 SDI survey data for Kenya to the country's 2014 Demographic and Health Survey (DHS) data is useful for examining the relationship between facility-level characteristics and a proxy for household wealth.

The DHS 2014 includes 36,430 households, each with an estimated wealth index constructed using principal components analysis on household asset data. These data are matched to SDI data on facility characteristics in each county. Analysis

is shown for all 47 counties in Kenya, with both the DHS 2014 and the SDI data representative at that level. The results show relatively even distribution of facility characteristics across the wealth index (figure B2.2.1). The only significant relationship is for the availability of health facility infrastructure, which is lower in counties with a lower wealth index. However, this relationship may be driven by the greater share of poor households in rural areas, because the relationship does not remain significant when controlling for the share of DHS households that are rural.

In 2013 Kenya devolved responsibilities for health care provision to counties, with the goal of increasing local accountability for the quality of services (McCollum et al. 2018). This descriptive analysis suggests that facility characteristics remain fairly equal across counties following the devolution, with the possible exception of infrastructure availability, which remains lower in poorer and more rural areas.

FIGURE B2.2.1   DHS wealth index and health facility-level characteristics in Kenya, by county, 2018
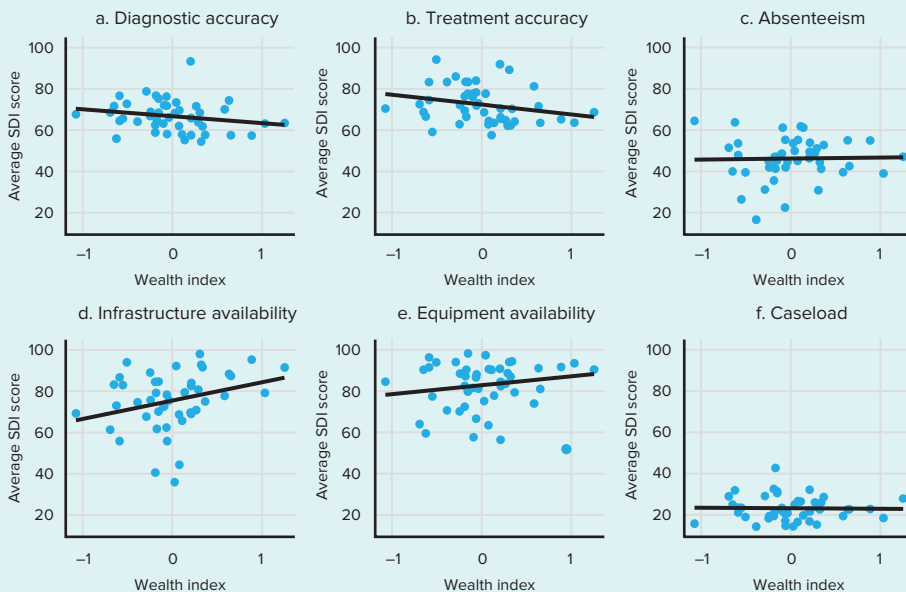
Source: Demographic and Health Survey (DHS), 2014; SDI (Service Delivery Indicators) health survey.
Note: Each dot represents a county, with the average county-level wealth index score on the x-axis. Results for six major SDI indicators are shown, with the county-level average SDI score on the y-axis. Solid black line represents the linear fit.

a subset of the WHO's Model List of Essential Medicines, and availability is calculated as the percentage of those 14 medicines that are in stock and unexpired on the day of the facility visit (WHO 2019). Are all facilities equipped with these essential pieces of equipment and common medicines?

The SDI surveys indicate that 67 percent of public and 79 percent of private facilities have all four pieces of equipment (that is, thermometer, stethoscope, blood pressure cuff, and weighing scale). Public facilities have more basic equipment available in Mozambique and Togo, whereas private facilities are better stocked in the remaining countries. Among public facilities, a slightly higher proportion of urban facilities (76 percent) than rural facilities (66 percent) has all four pieces of equipment. Although highly variable across the nine countries, there is no clear relationship between the availability of these basic tools and country-level living standards as proxied by average GDP per capita. Figure 2.11 shows the key pieces of equipment missing in different countries. Often, facilities are missing only one of the four necessary pieces of equipment, but more than 30 percent of facilities in Niger and Nigeria lack multiple items. Personal protective equipment is another important set of equipment for health facilities and is profiled in box 2.3 for Sierra Leone.

Medicine availability is notably lower at health posts (32 percent) and health clinics (46 percent) than at hospitals (65 percent). This pattern might be expected, but, considering that the list of medicines is fairly basic, they should ideally be available at all levels. Medicines are also less available at public

FIGURE 2.11 Availability of equipment in public health facilities in nine African countries, by country and type of equipment
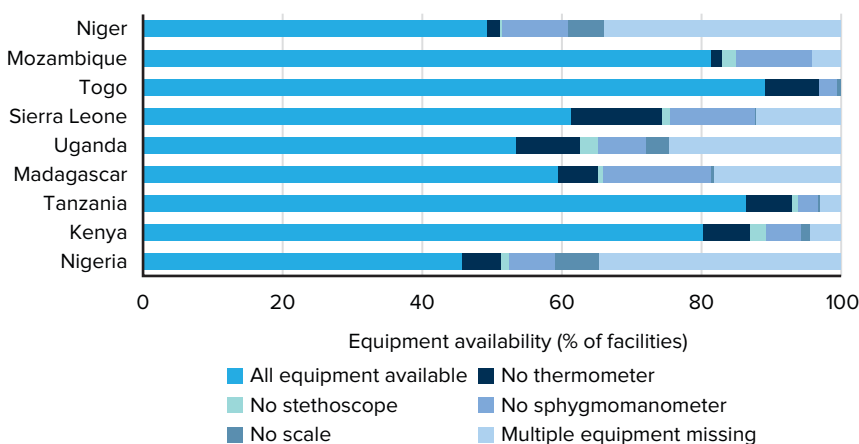
Source: SDI (Service Delivery Indicators) health surveys.
Note: "Multiple equipment missing" refers to facilities lacking more than one piece of equipment. Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013).

BOX 2.3   Pandemic preparedness in Sierra Leone

The COVID-19 (coronavirus) pandemic has called attention to systems' ability to protect health workers from acutely infectious disease. Personal protective equipment is used to ensure that health care workers are protected when serving high-risk patients. The SDI survey for Sierra Leone took place in 2018, following the 2013–16 West African Ebola epidemic, and collected data on the availability of disposable gloves, aprons, face masks, and protective boots, although information on protective eyewear and face shields was not available. Encouragingly, disposable gloves, face masks, and aprons were each available in more than 95 percent of facilities. Protective boots were less common, found in just over 70 percent of facilities. All four pieces of protective equipment were present in 66 percent of facilities.

Figure B2.3.1 shows the availability of these items for hospitals and health clinics and health posts. The high rates of ownership of personal protective equipment may be attributable partly to the Ebola response. Importantly, the SDI surveys (like other large-scale health surveys) do not collect information on the depth of stocks of equipment and supplies, meaning that a facility will be recorded as having disposable gloves even if only one pair is available. Given the breadth of the survey, counting stocks of supplies is not feasible; however, it also means that the estimates of availability described here do not necessarily reflect adequate supply. In the future, SDI surveys will include more information on pandemic preparedness (see chapter 4).

FIGURE B2.3.1   Availability of personal protective equipment in health facilities in Sierra Leone, by type of equipment and facility, 2018
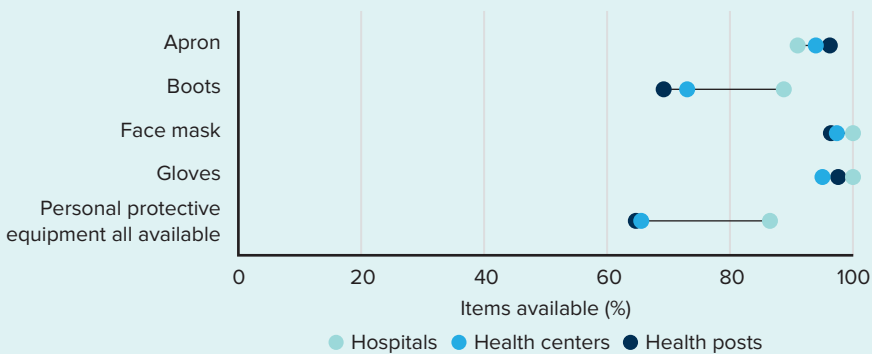
Source: SDI (Service Delivery Indicators) health surveys.
Note: The dots show discrepancies by type of facility.

facilities (36 percent) than at private facilities (54 percent). The difference for private and public facilities varies by country, from a 6-percentage-point difference in Togo to a 29-percentage-point difference in Tanzania, as shown in figure 2.12. Medicines are about equally available in rural and urban areas, suggesting that the location of facilities does not influence their stock of supplies. Availability of these medicines is low overall, with only 20 percent of facilities having more than half the medicines and less than 1 percent having all 14 medicines. Of note, some medicines, such as oxytocin, require refrigeration to

FIGURE 2.12 Percentage of health facilities with supplies of 14 basic medicines in six African countries, by country and type of facility

Medicine availability (% of essential medicines)

● Public average ● Country average ● Private average
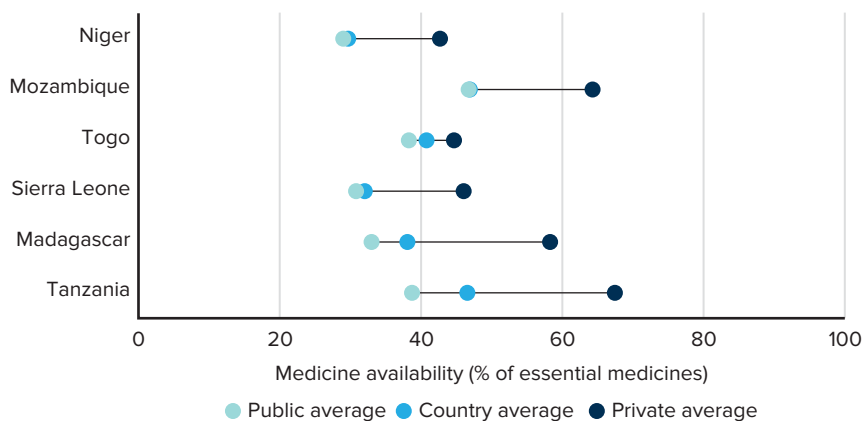
Source: SDI (Service Delivery Indicators) health surveys.
Note: Medicine availability is calculated as the share of 14 essential medicines available at a facility. The light green dots represent the average score at rural facilities, black dots represent the average score at urban facilities, and blue dots indicate the whole-country estimate. Country surveys were conducted in the following years: Madagascar (2016), Mozambique (2014), Niger (2015), Sierra Leone (2018), Tanzania (2016), and Togo (2013).

prevent spoilage, and only two-thirds of facilities (66 percent) report having a functional refrigerator.

Sterilization and waste disposal

Patients seeking care would also hope that the facility follows best practices with regard to equipment hygiene and cleaning procedures. A patient would hope to be treated with clean equipment that has been sterilized and protected from avoidable infections. Measures to promote prevention and control of infections are always important but especially so during a pandemic. Hand hygiene and appropriate waste segregation and disposal, for instance, are necessary for the safety of both patients and health care workers. The necessity of appropriate water, sanitation, and hygiene techniques was demonstrated during the 2014 Ebola epidemic in Sierra Leone, where the rate of infection was several times higher among health care workers than among the general population, partly owing to lack of robust infection prevention and control measures in health facilities (Kilmarx et al. 2014). How do health facilities fare in terms of waste disposal and sterilization?

Appropriate waste segregation and disposal practices are higher at hospitals (88 percent) than at health clinics (77 percent) or health posts (76 percent) and are also higher in urban areas and at private facilities. Similarly, equipment for

FIGURE 2.13 Availability of acceptable waste disposal and sterilization in public health facilities in seven African countries, by country
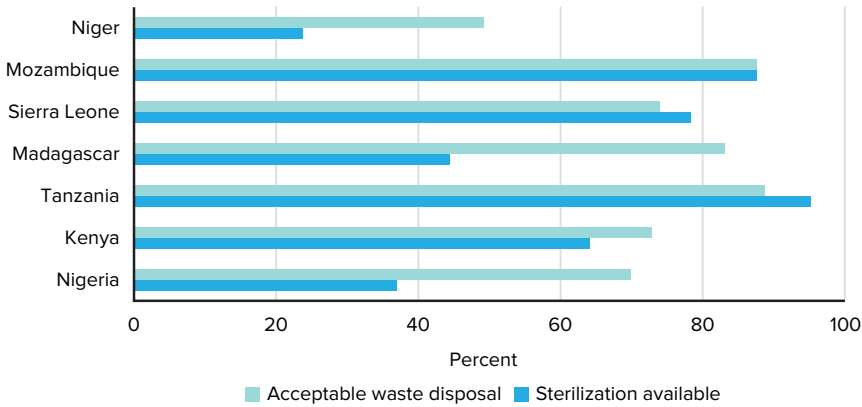
Source: SDI (Service Delivery Indicators) health surveys.
Note: The definition of acceptable waste disposal and sterilization is based on World Health Organization standards (WHO 2014). Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Sierra Leone (2018), and Tanzania (2016).

sterilization is more widely available at hospitals (91 percent) than at health clinics (75 percent) and health posts (51 percent) and is more available in urban areas and at private facilities. At the country level, Niger scores the lowest for both indicators, and Tanzania scores the highest. Countries tend to score higher on waste disposal than on sterilization, as shown in figure 2.13.

Can an average citizen's experience with the health system be improved?

The analysis presented in this section describes individual facets of service delivery. However, the experience of a typical patient is the result of a combination of facility- and provider-specific characteristics. Building on Di Giorgio et al. (2020), the following discussion assesses the availability of a combination of indicators to describe overall readiness to provide care. This analysis is intended to demonstrate how the wealth of information in SDI surveys can be recombined in novel ways to reveal system limitations.

Figure 2.14 depicts the availability of tools for diagnosis, the knowledge of health care providers in diagnosis and treatment, and the availability of medicines to treat each of the five health conditions studied by SDI. A set of appropriate equipment and medicines (as measured in the SDI survey) is selected for

FIGURE 2.14   Availability of key inputs for diagnosis and treatment of five common conditions in nine African countries
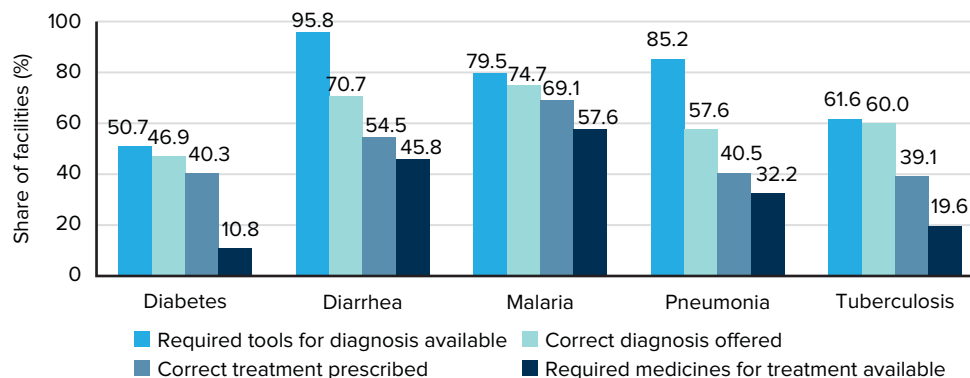
Source: SDI (Service Delivery Indicators) health surveys.
Note: Equipment for diagnosis is defined as a scale for diarrhea, a thermometer and rapid diagnostic test for malaria, and a stethoscope and thermometer for pneumonia. Appropriate equipment for diabetes and tuberculosis is not included in the SDI surveys, and therefore health care providers' mention of equipment in the vignettes is used as a proxy. For diabetes, appropriate equipment is assumed to be available if at least one health care provider at the facility ordered a fasting blood sugar, random blood sugar, glycated hemoglobin, or urinalysis test. For tuberculosis, appropriate equipment is assumed to be available if at least one health care provider ordered a chest X-ray, sputum microscopy, erythrocyte sedimentation rate, or C-reactive protein test. In general, these pieces of equipment are not intended to be exhaustive, but they are a selection of relevant equipment. Medicines for treatment are defined as hypoglycemics or insulin for diabetes, oral rehydration salts for diarrhea, artemisinin combination therapy for malaria, amoxicillin for pneumonia, and tuberculosis combination therapy for tuberculosis. Some countries are excluded due to lack of information. Kenya, Nigeria, Togo, and Uganda are excluded from the diabetes section because they did not collect information on diabetes medicines. Nigeria, Togo, and Uganda are excluded from the diarrhea section because their clinical vignettes included questions only on diarrhea with dehydration and not diarrhea alone. Kenya and Nigeria are excluded from the malaria section because they did not collect information on malaria diagnosis and treatment or malaria medicines. Kenya, Nigeria, Togo, and Uganda are excluded from the tuberculosis section because they did not collect information on tuberculosis combination therapy. Country surveys were conducted in the following years: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013).

each condition, and correct diagnosis and treatment refer to having at least one health care provider in the facility who is able to give correct answers on the related vignette. Each bar is conditional on the availability of inputs in the prior step, with the final bar representing the overall likelihood of a patient receiving all of the necessary steps in the care process for that specific ailment. For example, to treat a case of malaria, the required tools for diagnosis are a thermometer and a malaria rapid diagnostic test, the facility needs to have at least one provider who can accurately diagnose and treat malaria in the clinical vignette, and the required medicine for treatment is artemisinin combination therapy. Although most facilities have some of these individual components, only a little over half of facilities have all of the necessary components in combination and can therefore be considered prepared to treat a case of malaria.

Readiness to provide care differs across conditions, with a high of 57.6 percent of facilities prepared to provide care for a malaria patient and a low of 10.8 percent of facilities prepared to provide care for a diabetes patient. For diabetes and tuberculosis, a limiting factor is the lack of necessary tools and medicines. Lack of timely screening and diagnosis has been identified as a pressing issue for both of these diseases, and the results of this analysis suggest that primary care facilities still do not have the tools to address this problem or to provide appropriate medicines (Manne-Goehler et al. 2019; Raviglione et al. 2012). For the other conditions, no single factor emerges as dominant; rather, a combination of deficiencies results in facilities often being unprepared to offer full care.

Conclusions: What will it take to improve service delivery in health?

The SDI health surveys give insight into ordinary people's experience of PHC in nine Sub-Saharan African countries. SDI data shed light on the obstacles people encounter in seeking quality care for common medical conditions within these health systems and identify entry points for policy to improve PHC delivery and results.

Despite decades of global efforts to promote robust PHC, SDI evidence suggests that the quality of PHC delivery in these nine countries remains suboptimal. Upon arriving at a typical health facility, patients in these countries are likely to find a substantial number of clinical personnel absent. Despite the absences, many providers' outpatient caseloads are not especially elevated. This raises questions about how health systems organize and distribute their human resources. When health care providers are available, patients have a high likelihood of receiving an incorrect diagnosis and insufficient treatment. These risks

are especially pronounced at lower-level facilities where people typically make first contact with the health system.

Even if health care providers prescribe appropriate therapies, recommended essential medicines may not be available. The SDI surveys show that large numbers of health facilities still lack the basic infrastructure (electricity, water, sanitation), medical equipment, and sterilization facilities needed to provide quality PHC that respects patient safety. Importantly, SDI survey findings suggest substantial heterogeneity in the quality of PHC delivery between and especially within countries. An average citizen's experience with PHC, across these nine countries, depends to a large extent on where she is accessing care—whether in a rural or urban setting, at a public or private health facility, and at which level of facility.

What can be done to improve the average person's PHC experience in these health systems? In the aftermath of the COVID-19 pandemic, a burgeoning body of literature has noted the importance of strengthening the delivery of primary care services. The service delivery agenda moving forward should entail both an expansion and a reorganization of care to manage immediate risk and address long-term challenges simultaneously (World Bank 2020a). The results from the analysis presented in this chapter support the following directions for action.

- *Apply planning and management tools to reduce provider absence rates in the public sector.* Both unauthorized absences and total absence rates are higher in public facilities than in private ones, suggesting that practitioners in public facilities may have demands that take them outside of the facility, including attending trainings, providing outreach, and engaging in other authorized activities. Authorized absences, particularly in overburdened facilities, might reflect insufficient staffing or planning, speaking to the importance of better understanding staffing decisions and constraints and the role of management quality. Addressing absence rates is an important entry point for policy, because differences in provider absence rates between public and private facilities have the potential to widen health disparities further.

- *Rebalance caseloads and resources systemwide.* As noted, low outpatient caseloads might raise concerns about the effective allocation of human resources in health care delivery, but they also point to the success of global efforts to bring health facilities and health care providers closer to people, especially in rural and remote areas (WHO and World Bank 2017). Although expanded geographic access to care is important, especially from an equity perspective, low caseloads in primary care facilities across countries provide further evidence that services within the existing health system could be reorganized to enhance efficiency without compromising on equitable access, as proposed by the Lancet Global Health Commission on High Quality Health Systems (Kruk, Gage, Arsenault et al. 2018).

Quality-focused redesign of service delivery would entail treating chronic and stable conditions, preventive care, low acuity services, and palliative care at the primary level, while managing more complex or rare conditions in tertiary or specialized care centers.

- *Reinforce competencies among nurses and other less specialized cadres of health care providers in frontline facilities.* In terms of providers' diagnostic accuracy and therapeutic decisions, the SDI vignettes focus on common conditions that practitioners at all levels of the health care system should be able to diagnose and treat successfully. Diagnostic and treatment accuracy is higher at hospitals than at health centers or health posts. Because the majority of patients are likely to initially visit lower-level facilities, strengthening the capacity of of frontline workers is particularly crucial to ensure that they deliver quality primary care. Nurses and less specialized health providers make up the majority of the health workforce and need to be relied on for patient care. Yet they perform significantly worse than doctors on diagnostic and treatment accuracy as measured by the SDI vignettes. Improving these competencies likely requires improving both the quality of clinical education and the existing curriculum, beyond the current standards for in-service training. On the positive side, results from the vignette evaluations are not significantly different between health care providers at public versus private facilities or urban versus rural facilities, suggesting an encouraging equity in the current distribution of skilled health care providers.

- *Continue to improve supply chain management practices in the public system.* Key medicines and supplies are more commonly available in private as opposed to public facilities. The gap between public and private facilities persists even in urban areas, suggesting that it is not driven by the inaccessibility or remoteness of facilities. Further investigation is required to understand better the potential incentive structures, efficiencies, and management practices that may be driving better supply chain management in private facilities, so that these practices might be emulated in public ones.

- *Tackle infrastructure gaps at rural health facilities.* Shortfalls in medical equipment and basic infrastructure at many facilities pose urgent challenges for quality in PHC. Although the availability of basic infrastructure (improved water and sanitation facilities and electricity) is variable across countries in the sample, the starkest within-country contrasts are between urban and rural facilities. The implications of poor infrastructure at health facilities are dire: without safe water and sanitation, health care staff and patients are at increased risk of infection and associated illness (Sharma et al. 2020; WHO and UNICEF 2019). Facilities without access to electricity cannot operate crucial medical devices for

essential services (such as mammograms and electrocardiograms) and cannot reliably maintain a cold chain for storing vaccines or other medicines (Adair-Rohani et al. 2013). Scaling up solar power infrastructure at health facilities may help to address some of these cold chain challenges. The capacity for cold chain maintenance is taking on added importance now, as the global community expands provision of COVID-19 vaccines (Fischetti 2020). Given that rural health facilities may also be more likely to provide care to lower-income members of the population, poorer-quality infrastructure may contribute to greater health disparities by income, raising the stakes for policy action in this area.

Lessons from the SDI health surveys have already informed project design and supported health system reforms in some countries. Now, an exceptional window of opportunity exists for countries and development partners to accelerate these efforts. In the wake of COVID-19, public understanding of the importance of well-resourced health systems in saving lives and protecting economies is at an all-time high. In many settings, grassroots demand and high-level political buy-in for investment in resilient health systems are converging. Accordingly, COVID-19 has catalyzed long-overdue efforts to redesign health systems. SDI survey data on the quality of health service delivery can inform these ramped-up redesign processes both within and beyond Sub-Saharan Africa at a time when health systems are more stressed and more essential than ever.

Notes

1 | Details of sampling are discussed in appendix B.

2 | Five surveys are excluded, either because they were pilots (Senegal 2010 and Tanzania 2010) or because they were not yet complete and data were not publicly available at the time of writing (Cameroon 2019, Guinea-Bissau 2017, and Malawi 2019).

3 | See chapter 4 for more on comparability over time and results from the SDI surveys.

4 | The SDI survey on which the write-up for Nigeria is based was carried out eight years ago and in only 12 of the 36 states. Thus, it is not nationally representative, and circumstances may have changed since then.

5 | Further details are presented in appendix A, table A.1, and in appendix B.

6 | Sample weights at the facility level are not available for Mozambique, so unweighted results are reported. Provider-level weights are used for calculating the absence rate and for all measures related to the clinical vignettes.

7 | The estimates of GDP per capita (based on purchasing power parity in current international dollars) come from World Bank Open Data, and the year of the survey is used for each country's estimate. See https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD

8 | Further details on the calculation of each indicator are available in appendix A, table A.3.

9 | The category "other health workers" includes a variety of positions, depending on the country. It may include health assistants, community health workers, and midwives. In some contexts, it can include technicians, pharmacists, nutritionists, or orderlies, if they are involved with patient diagnosis and treatment.

10 | Diarrhea at 7.6 percent of disability-adjusted life years, malaria at 7.9 percent, diabetes at 1.3 percent, lower respiratory infections at 8.7 percent, tuberculosis at 3.3 percent, and anemia at 1.4 percent, based on IHME (2020).

11 | Correct treatment is not made conditional on correct diagnosis, so providers can occasionally prescribe the correct treatment without the correct diagnosis. Further information on the clinical vignettes is provided in Andrews et al. (2021).

12 | This section includes results from multivariate regressions presented in Andrews et al. (2021). Multivariate regressions for diagnostic and treatment accuracy include facility- and provider-level controls. The facility-level controls are ownership (public, private), location (urban, rural), facility level (hospital, health clinic, health post), and country. The provider-level controls are type, education, age (in 10-year groupings), and gender.

13 | In the diarrhea with dehydration vignette, the child presents as a case of diarrhea but displays multiple warning signs for severe dehydration. According to IMCI guidelines, severe dehydration necessitates rehydration with an intravenous line or nasogastric tube. However, 86 percent of providers simply prescribe oral rehydration salts (ORS), and 45 percent prescribe ORS plus zinc. ORS plus zinc is scored as appropriate treatment, because the child is able to drink in most vignettes. Similarly, for malaria with anemia, most providers identify malaria as the primary condition (diagnosed by 81 percent) but do not identify the warning signs for anemia (diagnosed by 21 percent) and therefore do not prescribe iron supplements.

14 | A facility is considered to have "improved toilets" if the enumerators confirms it has one or more functioning flush toilets or ventilated improved pit latrines, or covered pit latrine (with slab). A facility is considered as having an "improved water source" if it reports that its main source of water is piped into the facility, piped onto facility grounds, or comes from a public tap/standpipe, tube well/borehole, a protected dug well, a protected spring, bottled water, or a tanker truck. This definition is based on the WHO/UNICEF Joint Monitoring Program for Water Supply, Sanitation, and Hygiene. See also appendix A, table A.3.

References

Adair-Rohani, H., K. Zukor, S. Bonjour, S. Wilburn, A. C. Kuesel, R. Hebert, and E. R. Fletcher. 2013. "Limited Electricity Access in Health Facilities of Sub-Saharan Africa: A Systematic Review of Data on Electricity Access, Sources, and Reliability." *Global Health: Science and Practice* 1 (2): 249–61. https://doi.org/10.9745/GHSP-D-13-00037.

Andrews, K., R. Conner, R. Gatti, and J. 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.

Bamgboye, E. A., and A. I. Adeleye. 1992. "Sickness Absenteeism in a Nigerian Teaching Hospital." *East African Medical Journal* 69 (8): 450–55.

Belita, A., P. Mbindyo, and M. English. 2013. "Absenteeism amongst Health Workers—Developing a Typology to Support Empiric Work in Low-Income Countries and Characterizing Reported Associations." *Human Resources for Health* 11 (34): PMCID 3721994. doi: 10.1186/1478-4491-11-34.

Chaudhury, N., and J. Hammer. 2004. "Ghost Doctors: Absenteeism in Rural Bangladeshi Health Facilities." *World Bank Economic Review* 18 (3): 423–41.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan, and F. H. Rogers. 2006. "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *Journal of Economic Perspectives* 20 (1): 91–116.

Das, J., and K. Leonard. 2006. "Using Vignettes to Measure the Quality of Health Care." World Bank, Washington, DC.

Di Giorgio, L., D. K. Evans, M. Lindelow, S. N. Nguyen, J. Svensson, W. Wane, and A. W. Tarneberg. 2020. "An Analysis of Clinical Knowledge, Absenteeism, and Availability of Resources for Maternal and Child Health: A Cross-Sectional Quality of Care Study in 10 African Countries." Policy Research Working Paper 9440, World Bank, Washington, DC.

Endler, M., T. Al-Haidari, C. Benedetto, S. Chowdhury, J. Christilaw, F. El Kak, D. Galimberti, et al. 2021. "How the Coronavirus Disease 2019 Pandemic Is Impacting Sexual and Reproductive Health and Rights and Response: Results from a Global Survey of Providers, Researchers, and Policy-makers." *Acta Obstetricia et Gynecologica Scandinavica* 100 (4): 471–78. doi: 10.1111/aogs.14043.

Escamilla, V., L. Calhoun, J. Winston, and I. S. Speizer. 2018. "The Role of Distance and Quality on Facility Selection for Maternal and Child Health Services in Urban Kenya." *Journal of Urban Health* 95 (1): 1–12. https://doi.org/10.1007/s11524-017-0212-8.

Fischetti, M. 2020. "The COVID Cold Chain: How a Vaccine Will Get to You." *Scientific American,* November 19, 2020. https://www.scientificamerican.com/article/the-covid-cold-chain-how-a-vaccine-will-get-to-you/.

Hawkins, Merritt. 2012. "A Survey of America's Physicians: Practice Patterns and Perspectives." Survey conducted on behalf of the Physician's Foundation, Boston, MA.

IHME (Institute for Health Metrics and Evaluation). 2020. "Results from the Global Burden of Disease (GBD) Study." GBD Compare tool, Institute for Health Metrics and Evaluations, Seattle, WA. https://vizhub.healthdata.org/gbd-compare/.

Kaplan, R. S., M. Witkowski, M. Abbott, A. B. Guzman, L. D. Higgins, J. G. Meara, E. Padden, et al. 2014. "Using Time-Driven Activity-Based Costing to Identify Value Improvement Opportunities in Healthcare." *Journal of Healthcare Management* 59 (6): 399–412.

Kilmarx, P. H., K. R. Clarke, P. M. Dietz, M. J. Hamel, F. Husain, J. D. McFadden, B. J. Park, et al. 2014. "Ebola Virus Disease in Health Care Workers—Sierra Leone." *Morbidity and Mortality Weekly Report* 63 (49): 1168.

Kruk, M. E., A. D. Gage, C. Arsenault, K. Jordan, H. H. Leslie, S. Roder-DeWan, O. Adeyi, et al. 2018. "High-Quality Health Systems in the Sustainable Development Goals Era: Time for a Revolution." *Lancet Global Health* 6 (11): e1196–e1252. https://doi.org/10.1016/S2214-109X(18)30386-3.

Kruk, M. E., A. Gage, N. Joseph, G. Danaei, S. García-Saisó, and J. Salomon. 2018. "Mortality due to Low-Quality Health Systems in the Universal Health Coverage Era: A Systematic Analysis of Amenable Deaths in 137 Countries." *Lancet* 392 (10160): 2203–12.

Laxminarayan, R., A. Duse, C. Wattal, A. K. Zaidi, H. F. Wertheim, N. Sumpradit, E. Vlieghe, et al. 2013. "Antibiotic Resistance—The Need for Global Solutions." *Lancet Infectious Diseases* 13 (12): 1057–98.

Leslie, H. H., Z. Sun, and M. E. Kruk. 2017. "Association between Infrastructure and Observed Quality of Care in 4 Healthcare Services: A Cross-Sectional Study of 4,300 Facilities in 8 Countries." *PLoS Medicine* 14 (12): e1002464.

Liu, J., Y. Goryakin, A. Maeda, T. Bruckner, and R. Scheffler. 2016. "Global Health Workforce Labor Market Projections for 2030." Policy Research Working Paper 7790, World Bank, Washington, DC.

Lopez, C., A. Sautmann, and S. Schaner. 2021. "Does Patient Demand Contribute to the Overuse of Prescription Drugs?" Forthcoming in *American Economic Journal: Applied Economics.*

Manne-Goehler, J., P. Geldsetzer, K. Agoudavi, G. Andall-Brereton, K. K. Aryal, B. W. Bicaba, P. Bovet, et al. 2019. "Health System Performance for People with Diabetes in 28 Low- and Middle-Income Countries: A Cross-Sectional Study of Nationally Representative Surveys." *PLoS Medicine* 16 (3): p.e1002751.

McCollum, R., S. Theobald, L. Otiso, T. Martineau, R. Karuga, E. Barasa, S. Molyneux, and M. Taegtmeyer. 2018. "Priority Setting for Health in the Context of Devolution in Kenya: Implications for Health Equity and Community-Based Primary Care." *Health Policy and Planning* 33 (6): 729–42. https://doi.org/10.1093/heapol/czy043.

Parpia, A. S., M. L. Ndeffo-Mbah, N. S. Wenzel, and A. P. Galvani. 2016. "Effects of Response to 2014–2015 Ebola Outbreak on Deaths from Malaria, HIV/AIDS, and Tuberculosis, West Africa." *Emerging Infectious Diseases* 22 (3): 433–52.

Peabody, J. W., J. Luck, P. Glassman, T. R. Dresselhaus, and M. Lee. 2000. "Comparison of Vignettes, Standardized Patients, and Chart Abstraction." *Journal of the American Medical Association* 283 (13): 1715–22.

PHCPI (Primary Health Care Performance Initiative) and World Bank. 2020. "Key Messages: COVID-19 and Primary Health Care." PHCPI and World Bank, Washington, DC.

Rao, K. D., and A. Sheffel. 2018. "Quality of Clinical Care and Bypassing of Primary Health Centers in India." *Social Science and Medicine* 207 (June): 80–88.

Raviglione, M., B. Marais, K. Floyd, K. Lönnroth, H. Getahun, G. B. Migliori, A. D. Harries, et al. 2012. "Scaling up Interventions to Achieve Global Tuberculosis Control: Progress and New Developments." *Lancet* 379 (9829): 1902–13.

Roberton, T., E. Carter, V. Chou, A. Stegmuller, B. Jackson, Y. Tam, T. Sawadogo-Lewis, and N. Walker. 2020. "Early Estimates of the Indirect Effects of the Coronavirus Pandemic on Maternal and Child Mortality in Low- and Middle-Income Countries." *Lancet Global Health* 8 (7): e901–e908. https://www.thelancet.com/action/showPdf?pii=S2214-109X%2820%2930229-1.

Sharma, J., K. Andrews, R. Conner, and R. Gatti. 2020. "What Can the Service Delivery Indicator Surveys Tell Us about COVID-19 Preparedness?" Policy Research Working Paper 9334, World Bank, Washington, DC.

Speakman, E. 2016. "Effective Interprofessional Education and Collaborative Practice in Nursing Education." White Paper and webinar, October 15, 2016, Wolters Kluwer, Alphen aan den Rijn, Netherlands. https://www.wolterskluwer.com/en/expert-insights/effective-interprofessional-education?

Tumlinson, K., M. W. Gichane, S. L. Curtis, and K. LeMasters. 2019. "Understanding Healthcare Provider Absenteeism in Kenya: A Qualitative Analysis." *BMC Health Services Research* 19 (1): 1–8.

UN (United Nations). 2011. *Millennium Development Goals Report 2011*. New York: UN, June 2011. https://www.refworld.org/docid/4e42118b2.html.

UN (United Nations) General Assembly. 2015. "Transforming our World: The 2030 Agenda for Sustainable Development." Resolution A/RES/70/1, October 21, 2015, UN General Assembly, New York. https://www.refworld.org/docid/57b6e3e44.html.

UNICEF (United Nations Children's Fund). 2019. "WASH in Health Care Facilities: Global Baseline Report 2019." UNICEF, Paris.

UNICEF (United Nations Children's Fund) and WHO (World Health Organization). 1978. "Declaration of Alma Ata: International Conference on Primary Health Care, Alma Ata, USSR, 6–12 September 1978." WHO, Geneva.

Vermeersch, C., and P. I. Giovagnoli. 2020. "Eficiencia del gasto público en salud en Ecuador: Situación actual y opciones de política." World Bank, Washington, DC.

WHO (World Health Organization). 2006. *World Health Report 2006: Working Together for Health*. Geneva: WHO.

WHO (World Health Organization). 2008. *World Health Report 2008: Primary Health Care: Now More Than Ever*. Geneva: WHO.

WHO (World Health Organization). 2013a. "Arguing for Universal Health Coverage." WHO, Geneva. https://apps.who.int/iris/handle/10665/204355.

WHO (World Health Organization). 2013b. *World Health Report 2013: Research for Universal Health Coverage*. Geneva: WHO.

WHO (World Health Organization). 2014. "Safe Management of Wastes from Health-Care Activities." WHO, Geneva.

WHO (World Health Organization). 2019. "Model List of Essential Medicines." WHO, Geneva.

WHO (World Health Organization). 2020. "Role of Primary Care in the COVID-19 Response." WHO Regional Office for the Western Pacific, Manila. https://apps.who.int/iris/handle/10665/331921.

WHO (World Health Organization), OECD (Organisation for Economic Co-operation and Development), and World Bank. 2018. *Delivering Quality Health Services: A Global Imperative for Universal Health Coverage*. Geneva: WHO, OECD, and World Bank.

WHO (World Health Organization) and UNICEF (United Nations Children's Fund). 2019. *Water, Sanitation, and Hygiene in Health Care Facilities: Global Baseline Report 2019*. Geneva: WHO and UNICEF.

WHO (World Health Organization) and World Bank. 2017. *Tracking Universal Health Coverage: 2017 Global Monitoring Report*. Geneva: WHO and World Bank.

World Bank. 2020a. "COVID-19 and Health Systems: Just the Beginning." World Bank, Washington, DC.

World Bank. 2020b. "From Data to Action: Big Data, GIS Tools, and Machine Learning to Modernize the Delivery of Health Services in Peru." World Bank, Washington, DC.

3. Education service delivery in nine African countries

Background: Reimagining what education can achieve

Education is important for empowering citizens, developing a skilled workforce, enabling upward socioeconomic mobility, improving economic growth, and fostering a prosperous society. Article 26 of the 1948 Universal Declaration of Human Rights recognized that everyone has the right to education, a principle that translated into the promotion and expansion of access to quality education for all through the Millennium Development Goals and Sustainable Development Goals (UN 2011; UN General Assembly 2015).

Before the COVID-19 (coronavirus) pandemic, most national education systems were already facing a crisis. Its nature was spelled out in the *World Development Report 2018: Learning to Realize Education's Promise* (World Bank 2018) and in the United Nations Educational, Scientific, and Cultural Organization's 2013 report *The Global Learning Crisis: Why Every Child*

Deserves a Quality Education (UNESCO 2013). Despite the expansion of access to schooling in recent decades, many low- and middle-income countries have not been able to translate increased enrollment numbers into higher levels of learning for their children. This shortfall is reflected in the troubling evidence provided by the Learning Poverty measure, according to which 53 percent of children 10 years old in the world today cannot read or comprehend a simple text (World Bank 2019).

Before the pandemic, the learning crisis had multiple roots. Unprepared learners, teachers with insufficient skills and motivation, scarce or deficient school inputs, poor school management, and weak governance all contributed. The result in many countries was poor-quality service provision and education systems that did not work for children. However, learning shortfalls did not affect all countries and all children equally. According to the 2016 International Commission on Financing Global Education Opportunity, 90 percent of children in low-income countries, compared to only 30 percent in high-income countries, fail to master basic secondary-level skills on time (International Commission on Financing Global Education Opportunity 2016). Structural differences associated with poverty, gender, ethnicity, disability, and location explain a substantial portion of schooling disparities (World Bank 2018). Worldwide, girls are twice as likely as boys never to start school (International Commission on Financing Global Education Opportunity 2016).

The COVID-19 pandemic has exacerbated these preexisting disparities. At the height of the pandemic in 2020, 1.6 billion children worldwide were not physically in school (UN 2020). Coming atop significant losses of family income, this unprecedented disruption of education systems has upended learning in many settings, with the potential to scar children's learning and school attainment for years—perhaps generations—to come (World Bank 2020). Given differences in access to digital devices, internet connectivity, parental involvement, time for supervision, and other factors, the disruption is likely to affect children from disadvantaged families most severely. In many countries, business closures and lockdowns are taking a heavy economic toll on families engaged in informal work. As a result, pressures are growing on many children in poor and vulnerable households to drop out of school temporarily or permanently. Understanding the bottlenecks to learning is a necessary step for the global education community to rethink and reinvigorate schooling in the face of this disruption.

SDI education surveys: Seeing basic education from the students' perspective

Faced with underlying structural shortfalls in learning, which are exacerbated by the effects of the COVID-19 shock, innovative solutions are needed to protect learning now and lay the foundations for more efficient, equitable, and resilient

systems tomorrow. This chapter contributes to the existing but still incomplete body of evidence in low- and middle-income countries; it is intended to inform decision-makers and implementers facing tough choices on how to "build back better" in education.

The Service Delivery Indicators (SDI) education surveys offer a set of indicators for benchmarking the quality of primary education being delivered. Unlike surveys for the health sector, the SDI education surveys directly measure a crucial human capital outcome: student learning for fourth-grade pupils. By assessing an outcome, SDI education surveys show how learning for children is related to key elements of the provision of education: teachers' knowledge, teachers' effort and time spent teaching, and availability of school infrastructure and materials needed to teach effectively.

As the SDI health surveys reflect the experience of a typical patient moving through the health care process, so the SDI education studies capture core features of a typical student's experience of school and the results that her schooling enables her to achieve. For children to learn, teachers need to be present and to know the material beforehand. Likewise, students need to have paper and pen as well as textbooks. The presence of a blackboard that is visible to all students is essential, as is the availability of basic infrastructure. SDI surveys collect information on all of these variables from a school administrator and through direct observation and relate these findings to how much children are actually learning.

To reflect the progress and pitfalls of students' learning journey, this chapter draws on a decade of SDI education data spanning nine countries. The chapter begins by reviewing the methodology of the SDI education surveys and the types of data obtained. The chapter's core sections then set out and analyze SDI survey findings, highlighting differences in learning outcomes between and within countries and organizing evidence to explain these differences. By comparing key characteristics of high- and low-performing schools, the chapter identifies promising levers that policy makers may use to improve outcomes and reduce disparities in educational achievement, fulfilling the promise to build back better. Because private education is difficult to compare across countries, the bulk of the analysis focuses on public schools. A complementary analysis in the concluding sections also incorporates private schools in a subset of SDI countries.

Sample, methods, and framework

Over the past decade, the SDI program has collected data about schools and how learning happens in nine African countries (table 3.1). The program has surveyed 3,297 schools, collecting information from more than 35,000 teachers and 32,000 students on school-level characteristics; teachers' effort, knowledge, and pedagogy; and learning outcomes. The information extracted from

TABLE 3.1 Size of the SDI education sample in nine African countries, by country

Country	Survey year	Schools	Teachers	Pupils
Kenya	2012	306	4,425	2,952
Madagascar	2016	473	3,049	3,970
Morocco	2016	299	3,052	2,917
Mozambique	2014	203	1,950	2,030
Niger	2015	256	2,140	3,507
Nigeria	2013	760	6,146	6,735
Tanzania	2016	400	6,979	4,825
Togo	2013	200	1,238	1,938
Uganda	2013	400	6,073	3,963
All		3,297	35,052	32,837

Source: SDI (Service Delivery Indicators) education surveys.
Note: The number of teachers includes every teacher on the school roster, from which a subsample was tested, another subsample was selected for the absenteeism module, and a smaller proportion was observed in the classroom.

these comprehensive surveys is representative at the national level. It is also representative of rural and urban schools and of private and public schools.[1]

As in the health sector, the comparability of the SDI education surveys across countries enables reasonable benchmarking, which, in turn, highlights the salience of knowledge and infrastructure gaps to foster momentum for reform. This chapter incorporates data from the following country education surveys: Kenya (2012), Madagascar (2016), Morocco (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Tanzania (2016), Togo (2013), and Uganda (2013). The combination of these countries represents approximately 39 percent of the population of schoolchildren in Sub-Saharan Africa.[2] As in the previous chapter, for all figures, countries are ordered by increasing GDP per capita to explore how performance on education service indicators may correlate with country income.[3]

The school sample in each country is usually drawn from the national school census or a similar list facilitated by the government. The process leading to the survey is intended to build on a dialogue with the country's ministry of education, and questionnaires balance the global objective of comparability with country-specific priorities. Country sample sizes vary according to the size of the country and the level of representativeness intended, ranging from 200 schools in Togo, with representativeness at the national level, to 760 schools in Nigeria, with representativeness at the level of each of the four selected states (Anambra, Bauchi, Ekiti, and Niger). To date, on average, a typical SDI education survey includes just over 360 schools, which translates into approximately 3,650 students tested and more than 4,000 teachers observed and surveyed per country. For these nine African countries, the sampled schools tend to include more rural and public schools. Very few private schools were sampled in Morocco, Mozambique, Niger, and Tanzania, because private schools are only a small

fraction of the education system in those countries. Overall, about 70 percent of schools in the cross-country sample are public.

The SDI education surveys focus on eight indicators that measure fundamental conditions that shape ordinary students' school experience: rates of teacher absence from both the school and the classroom, time spent teaching or time on task, minimum teacher knowledge, availability of minimum infrastructure, availability of minimum classroom equipment, share of pupils with textbooks, and pupil-teacher ratio. Additionally, student learning is measured through a student assessment on math, language, and nonverbal skills. These indicators are intended to produce a comprehensive overview of the current state of each education system, putting learning at the center.

In order to collect observational and complete information, schools are visited twice. The first is an announced visit to collect facility-level data on inputs and a full roster of teachers and their qualifications. During this visit, up to 10 teachers per school currently teaching in the fourth grade are tested on math, language, and pedagogical knowledge. Additionally, one fourth-grade language or math lesson is randomly selected and observed with detailed note taking, and up to 10 students from that classroom are randomly selected and tested on math, language, and nonverbal reasoning. Both the teacher and student assessments are based on content derived from the third- and fourth-grade national curricula of 13 African countries.[4]

During a second, unannounced visit, the school and classroom presence or absence of up to 10 presampled teachers is recorded. Unlike available administrative data, SDI surveys base all indicators, including teacher absence, on direct observations rather than self-reported or supervisor-reported data. During the second visit, the enumerator directly counts unstaffed classrooms. These strategies mean that SDI data can closely approximate the daily realities of school as students actually see and feel them. In other words, SDI surveys provide a de facto picture of how social services work for the people they are meant to benefit.

Using the information from these two visits, SDI education surveys measure student learning and generate findings in three key areas:

1. *Provider's effort*. Are teachers present at school and in the classroom during their scheduled hours? Do teachers use the time to teach effectively during classroom observations?
2. *Provider's knowledge*. Do teachers have the minimum math, language, and pedagogical knowledge to teach effectively at the fourth-grade level?
3. *Inputs*. Do schools and classrooms have the necessary equipment (for example, a functioning blackboard, chalk, pens, pencils, and exercise books in fourth-grade classrooms) and infrastructure (for example, functioning toilets and classroom visibility) to deliver quality schooling? Are there enough teachers for students? Do students have textbooks to learn from?

The remainder of this chapter uses these questions to describe students' learning experience in nine countries. As noted, because private educational institutions

and practices vary widely across countries, the first sections of the chapter focus exclusively on public schools, which are more easily comparable. Later sections feature an analysis incorporating private schools for a subset of countries.

Harnessing SDI data to improve systems

The analysis here builds on previous work using SDI education data (Bashir et al. 2018; Bold et al. 2017; Bold et al. 2019; Mbiti 2016). Most notably, Bold et al. (2017) examine teacher absence, time spent teaching, teacher knowledge, and pedagogical skills in primary schools using data from eight SDI surveys.[5] The paper documents that, on average, 44 percent of teachers are absent from class or school when they are supposed to be present, with absence rates ranging between 23 percent (Nigeria) and 57 percent (Uganda). In three of the eight SDI surveys analyzed, more than half of teachers are absent from the classroom on a given day. Only two-thirds of teachers know at least 80 percent of the fourth-grade curriculum, with large variation across countries. More than 90 percent of teachers in Kenya and Uganda have mastered the knowledge that their students are supposed to learn, compared with only 25 percent of Nigerian teachers. The level of pedagogical skills is equally concerning. More recently, Bold et al. (2019) use SDI data to document that insufficiency in teachers' content knowledge accounts for 30 percent of the shortfall in learning relative to the curriculum and about 20 percent of the cross-country difference in learning in the sample.

This chapter builds on these previous findings along three important dimensions. First, it uses fully harmonized data for a larger set of countries. Second, it provides a more extensive description of the factors associated with learning—for instance, by analyzing differences in school infrastructure. Finally, it exploits the large number of schools surveyed within each country to document the large heterogeneity that exists across schools in terms of both learning and associated factors.

How much is the typical student learning?

SDI education surveys are designed to bring ordinary students' school experience and outcomes into clear focus in the countries that implement the studies. What characterizes that experience, in broad terms, across the nine SDI countries? The answer starts with the physical and institutional organization of the school itself. Across the SDI education sample, a typical primary school includes first through sixth grades, although certain schools also offer some preprimary levels. On average, each school has about 390 children in the primary level, and enrollment is distributed equally between girls and boys. There are about nine classrooms per school, with 47 children per classroom. Each school has 11 teachers, which translates into 36 students per teacher, on average. This ratio is

notably higher than in high-income countries, such as the United States, where the average student-to-teacher ratio in public schools is 16:1 (Hanson 2021).

A typical student in fourth grade in the SDI sample is about 11 years old and equally likely to be a boy or a girl, except in the four states surveyed in Nigeria, where the ratio of boys to girls is about 2:1. Students generally have breakfast at home (70 percent), but some have no breakfast (22 percent), and some receive breakfast at school (8 percent). When the typical student arrives at school, on one out of every three days she finds her teacher absent. When present, teachers are likely to be underprepared. The average teacher in the SDI sample answers correctly only 4 out of 10 language and math questions and 2 out of 10 pedagogy questions. The typical student has access to a textbook less than half of the time, although most classrooms have a functioning blackboard. Overall, the typical student and her peers learn, but not as much or as fast as they could. Fewer than half of students are able to read a simple sentence correctly by fourth grade.

Low student achievement overall

Cognitive skills play an important role in explaining educational attainment (Cameron and Heckman 2001; Heckman, Stixrud, and Urzua 2006), and acquiring them early makes it easier to improve them later (Cunha and Heckman 2007). There is strong evidence that verbal and math skills are particularly important for future academic and labor market outcomes. In other words, a student's ability to read and do basic math in fourth grade will likely affect many important outcomes in her life. Later sections of this chapter look in detail at features of the school context that previous literature (World Bank 2018) has found to be relevant for learning. The SDI surveys highlight a series of these factors.

On average, students in public schools show low levels of learning across all nine countries in the SDI sample (figure 3.1). When tested with the SDI student assessment, less than half of children can correctly read a simple sentence out loud, and less than 40 percent can correctly perform single-digit multiplications. Although more than 90 percent of children can identify three numbers, less than 15 percent can multiply triple digits. Children are expected to master these simple tasks by the end of fourth grade. When looking at differences by gender, there is almost no difference in language performance, although male students tend to do slightly better than female students in math.[6]

The typical student in the bottom decile of the distribution across all SDI countries is able to identify three numbers less than 50 percent of the time, identify the smallest among a group of fractions less than 8 percent of the time, and complete a sequence of numbers less than 4 percent of the time. This student cannot answer any other math question correctly, including single-digit additions, subtractions, multiplications, and divisions. Similarly, in the language assessment, the bottom-decile student can only identify three letters correctly 22 percent of the time and match words with pictures correctly 34 percent of the time. In this group of students, every other language question, including

FIGURE 3.1 Student performance in language and mathematics, average results for nine African countries
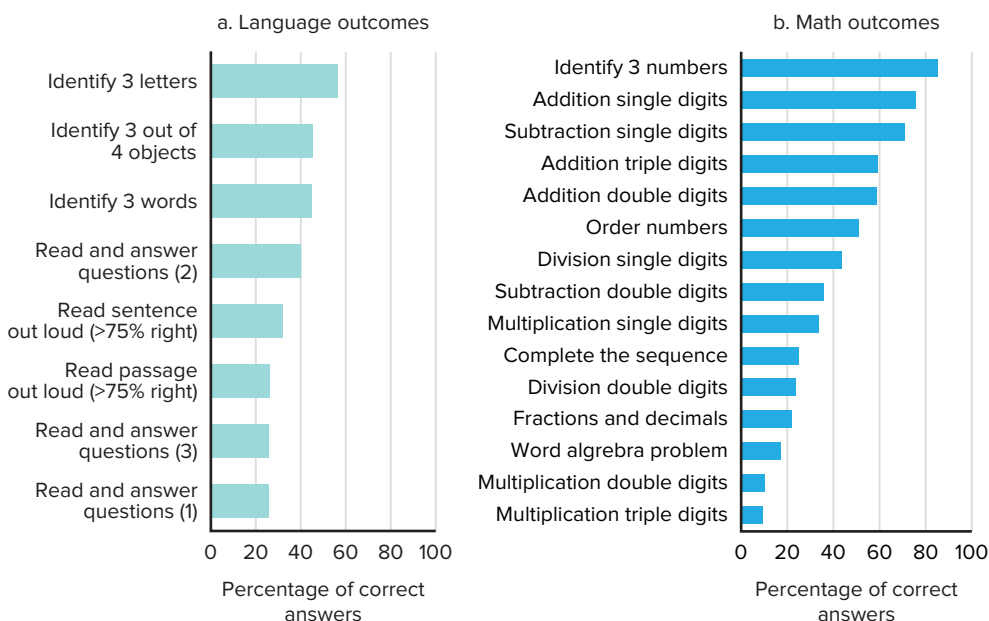
a. Language outcomes

b. Math outcomes

Percentage of correct answers

Percentage of correct answers

Source: SDI (Service Delivery Indicators) education surveys.
Note: This figure shows the average share of students who correctly answer each task or question on the language and mathematics SDI student assessments. The figures show the simple average across all students in public schools only. On panel a, the numbers in parentheses are the number of questions to answer. Country surveys were conducted in the following years: Kenya (2012), Madagascar (2016), Morocco (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Tanzania (2016), Togo (2013), and Uganda (2013).

reading a simple sentence, is answered correctly less than 10 percent of the time. The typical student at the bottom of the distribution is able to read a simple paragraph and correctly answer a basic question less than 8 percent of the time. Low reading proficiency also has implications for math knowledge and a snowballing effect on later learning (World Bank 2019).

Achievement gaps within and between countries

While the low average levels of learning are discouraging, the SDI surveys document substantial differences in student performance within and between countries (figure 3.2).[7] In Niger, for example, the gap between the highest- and lowest-performing students is particularly striking. Indeed, in almost every country some students are performing at the bottom of the overall distribution, whereas some students are obtaining perfect scores. This finding suggests that producing better-prepared students supported by schools that enable learning is not beyond the reach of any of these countries.

Differences in learning among countries are large,[8] with seemingly three clusters: Kenya, Madagascar, Morocco, and Tanzania at the top; Nigeria, Togo, and

FIGURE 3.2 Standardized performance on the SDI student assessment in nine African countries, by country

a. Student math test score distribution by country
Public schools

b. Student language test score distribution by country
Public schools

c. Mean standardized test scores versus GDP
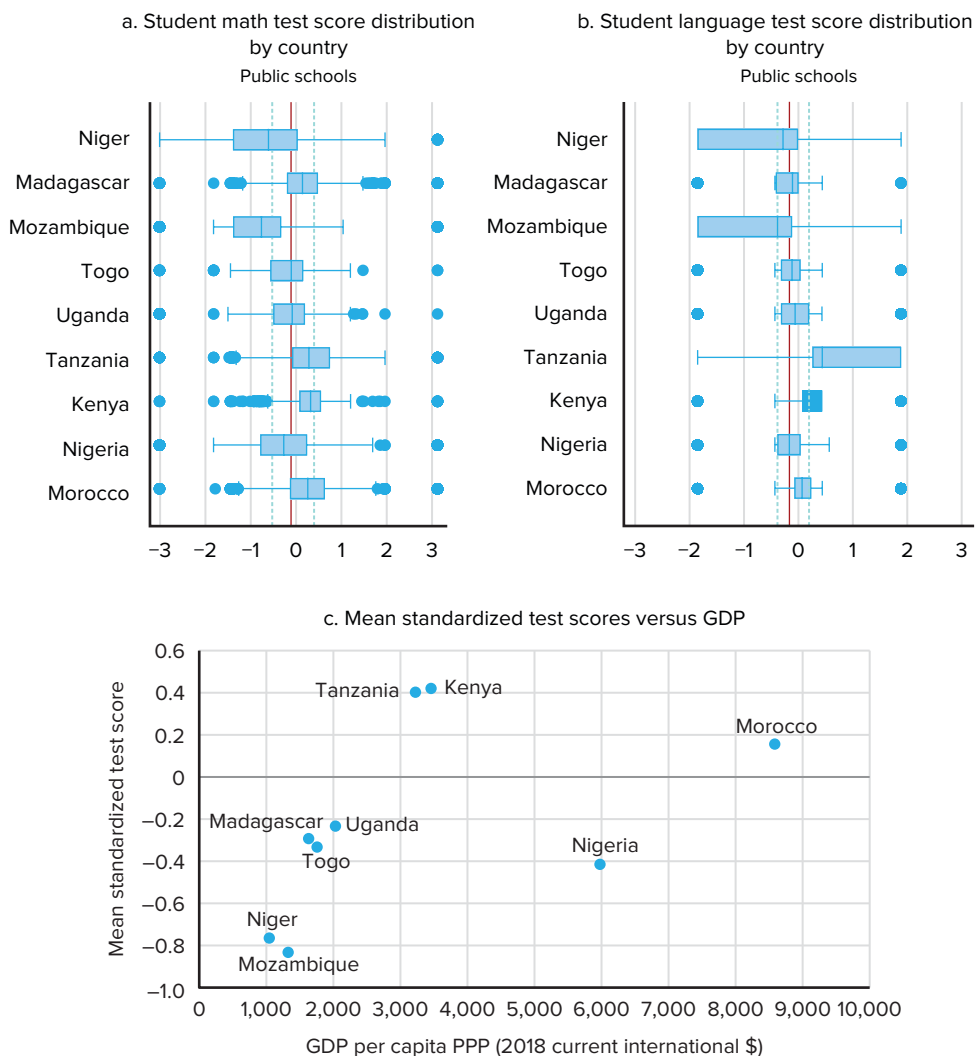
Source: SDI (Service Delivery Indicators) education surveys.
Note: This figure shows the standardized student test scores for public schools only. Panels a and b show standard boxplots displaying a line for the lower and upper adjacent values and a box with the median, 25th, and 75th percentiles for each country. The test scores are equated across countries using an IRT 2-PL model. The boxplots use MLE results. The solid line represents the mean for all countries, and light green patterned lines represent the 25th and 75th overall percentiles. Countries are ordered by GDP per capita PPP, from lowest to highest. Panel c shows a scatter plot between the mean standardized student test score combined for both language and mathematics and each country's GDP per capita PPP. Country surveys were conducted in the following years: Kenya (2012), Madagascar (2016), Morocco (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Tanzania (2016), Togo (2013), and Uganda (2013). IRT 2-PL = item response theory 2-parameter logistic; MLE = maximum likelihood estimation; GDP = gross domestic product; PPP = purchasing power parity.

Uganda in the middle range, and Mozambique and Niger at the bottom, with the lowest average scores. The mean difference between the countries with the highest SDI performance (Tanzania) and the lowest performance (Mozambique) is 1.2 standard deviations.[9]

To appreciate the size of this gap, one can think of two groups of students who differ by 1.2 standard deviations in their test scores: a high-scoring group and a low-scoring group. If a student were in the high-scoring group, she would master 8 out of 10 items. If she were in the low-scoring group, she would master only two items. The items that both groups would master are of low difficulty: identifying three numbers and adding single-digit numbers. However, the high-scoring group would also master more complex items, such as adding triple digits and dividing single digits, as shown in figure 3.3. Another way of understanding this difference is to compare it to the gap in Program for International Student Assessment test scores between two countries. A difference of 1.2 standard deviations is equivalent to the gap between Ecuador and the mean score for Organisation for Economic Co-operation and Development

FIGURE 3.3 Visualizing differences in test scores between high- and low-performing groups of students in nine African countries
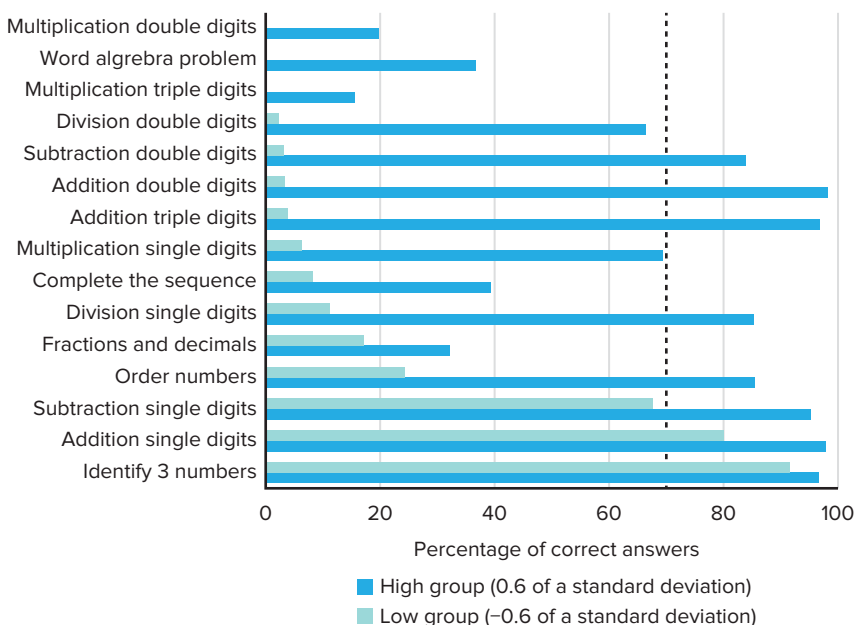
Source: SDI (Service Delivery Indicators) education surveys.
Note: This figure shows the mean proportion of correct answers per task or question for the SDI mathematics student assessment for public school students across all countries combined. Students performing around −0.6 and 0.6 of a standard deviation are included in the low- and high-performing groups, respectively, to illustrate the potential difference of 1.2 standard deviations. Results might vary if different groups were selected. Dashed black line = 70 percent (arbitrarily set proficiency). Country surveys were conducted in the following years: Kenya (2012), Madagascar (2016), Morocco (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Tanzania (2016), Togo (2013), and Uganda (2013).

countries (OECD 2018) or the difference between Singapore, one of the world's top-ranked countries in math, and Serbia, ranked 46 (OECD 2018).

Differences across and within countries coexist with large within-school differences in learning. A simple variance decomposition reveals that, across SDI countries, on average, about 25 percent of the variation in student test scores comes from between-school variation, which is in line with the Organisation for Economic Co-operation and Development (OECD) Program for International Student Assessment (PISA) average of 33.6 and 37 percent in 2004 and 2012, respectively (OECD 2004, 2013), as well as with the 34 percent estimated variation obtained using the World Bank Global Learning Assessment Database (GLAD) (Azevedo and Goldemberg 2020). Although it is possible for students to have different experiences in the same school because of factors such as teacher biases, the large amount of unexplained variation suggests that factors beyond the school level might also be influencing learning. Unfortunately, the information available on students' characteristics and home environment is limited in the SDI sample analyzed here.[10] More recent SDI surveys are attempting to fill this gap (see chapter 4). The language of instruction may also affect student scores, as detailed in box 3.1.

BOX 3.1 How does language of instruction affect test scores?

Students are normally tested in the official language of instruction, especially for international and national large-scale assessments. This practice works well in monolingual contexts but creates numerous problems in multilingual ones. Students for whom the language of instruction is not their mother tongue (L1) systematically score lower than students for whom it is.

The literature documenting lower scores is voluminous and consistent. Mullis et al. (2017) find that 92 percent of students from the 48 countries tested in the 2011 Progress in International Reading Literacy Study (PIRLS) assessment spoke the language of the test at home. Those not tested in their L1 scored significantly lower, by more than one-third of a standard deviation. Glewwe, Chen, and Katare (2012) find that linguistic-minority students fall behind very early in their school experience and have a hard time catching up. Other authors have documented the strong correlation between being taught in one's L1 and continuing in primary school (Ramachandran 2012).

Do these persistently lower scores indicate lower achievement, test bias, or both? If there

is test bias, is it due to poor translation or to deeper "configural problems," when the constructs themselves fundamentally differ between languages? Translation problems are relatively easy to spot and fix with enough resources, but configural problems pose a greater challenge.

At the heart of the problem is the fact that tests in a single language cannot distinguish students who answer incorrectly because they truly do not know the construct and those who could answer correctly if the question were asked in their mother tongue. When students are grouped and tested in their L1, analyses can, in theory, estimate the extent of bias between versions of the test. In practice, this way of testing and analyzing results is rare outside of the main languages spoken in Organisation for Economic Co-operation and Development (OECD) countries (Ramachandran 2012).

One part of the story is clear: the problem is not the inability of bilingual or multilingual students to achieve as well as or better than monolingual students. Collier and Thomas (2017) find that, when students receive enough high-quality instruction in both

(Continued)

BOX 3.1   How does language of instruction affect test scores? (*Continued*)

languages, bilinguals will eventually outperform monolinguals even in the monolinguals' L1. But instruction almost invariably falls short of optimal, and so knowing by how much a linguistic minority is lagging, and why, is critically important.

The problem is becoming more important as testing coverage expands globally. International large-scale assessments were initially designed for and first given in OECD member countries, which tend to be more linguistically homogeneous than non-OECD countries.[a] In 2000, the Program for International Assessment (PISA) had 41 national test versions in 25 languages for 30 participating (OECD member) countries; by 2006, 77 versions in 42 languages were given, with all of the increase from non-OECD member countries. The expansion "added considerably to the challenge of ensuring equivalence and fairness of instruments across all participating countries" (Grisay et al. 2007).

The challenge is formidable, but, by testing students in their L1 and appropriately analyzing differences between language groups, progress is possible. In fact, some initiatives are already moving in the right direction. For instance, the International Association for the Evaluation of Educational Achievement (IEA) has created guidelines for countries participating in PIRLS and other international large-scale assessments. Countries are responsible for translating the assessment into their own languages and adapting it to their own contexts. In the same spirit, IEA and Boston College conduct studies to detect test and item bias following standards in the field of psychometrics (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education 2014; Educational Testing Service 2014). In the instances where measurement bias is identified (due to language at home, gender, or other factors), these organizations are transparent in communicating these results.

The growth in participation in international large-scale assessments provides an opportunity for many countries and for international development organizations. Organizations that conduct international large-scale assessments support participating countries with capacity-building initiatives so that they can conduct better national large-scale assessments and follow best assessment practices.

Source: Contributed by Michael Crawford.
a. Ethnologue data, 22nd ed. (https://www.ethnologue.com).

Among the determinants of student learning, SDI surveys primarily collect information on school inputs and teacher characteristics. For that reason, the remainder of this chapter focuses principally on variations in student learning that can be explained by differences in these characteristics.[11]

Are basic requirements for learning in place?

Many factors—both internal and external to education systems—contribute to a student's ability to learn those basic skills that will stay with her or him throughout life.[12] While individual schools are affected by the broad characteristics of the country's education system and its stakeholders, factors at the school level decisively influence the learning experience of students. Describing some of these factors is the comparative advantage of surveys such as the SDI.

Are teachers present and teaching?

Teachers need to be present in class to teach. Not only has teacher absence been found to correlate with lower learning, but causal studies also have shown that reducing absence can improve learning (Duflo, Hanna, and Ryan 2012). However, even when they are in school, teachers often spend too much time on activities other than teaching. As mentioned earlier, teacher absence in SDI countries is well documented (Bold et al. 2017). The analysis conducted for this chapter, albeit in an updated sample of SDI countries and using a slightly different definition due to the careful harmonization of the surveys,[13] yields a similar story. On average, 22 percent of teachers are absent *from school* during a surprise visit. If teachers who are *not in the classroom* during this visit are also counted, the teacher absence rate rises to 38 percent.[14] Overall, teacher absence remains a substantial challenge for SDI countries included in this book. There are many possible reasons for teacher absence, including systemwide shortfalls in personnel policies (Liu, Loeb, and Shi 2020), lack of monitoring and accountability, and insufficient incentives (Mbiti 2016; Muralidharan et al. 2016).

Do teachers have the knowledge and skills they need?

The importance of teacher quality and, in particular, of effective pedagogy has been amply documented in the education literature (see Araujo et al. 2016; Evans and Popova 2016; Hanushek and Rivkin 2006). Teachers' abilities are often assumed to be associated with academic credentials. However, a growing body of evidence shows that the skills that matter most for learning—content knowledge and pedagogy—are not necessarily linked with teachers' formal qualifications (Cruz-Aguayo, Ibarrarán, and Schady 2017; Hanushek and Rivkin 2012; Rivkin, Hanushek, and Kain 2005). By providing direct measures of knowledge and pedagogy, SDI surveys make it possible to measure the importance of teachers' abilities in explaining children's learning outcomes. In fact, a recent study uses SDI surveys and other data to show that the associations between student test scores and teacher attributes might differ for teachers who have high and low scores on content knowledge and pedagogy (Filmer, Molina, and Wane 2020).

The SDI teacher assessment includes two sections. The teacher knowledge section resembles grading a math and literacy exam (grade arithmetic exercises solved by students, correct a letter with grammatical errors, and similar tasks), whereas the pedagogical section asks teachers to perform tasks that they face on a daily basis (prepare to teach a lesson, assess differences in children's abilities, and evaluate students' learning achievements and progress). The extracts from teachers' tests shown in figure 3.4 are examples of the types of questions on which teachers are assessed.

FIGURE 3.4 Extracts from teacher assessments

No.	Question	(A) Mark the pupil's answer correct or incorrect here	(B) Write the correct answers here
2	*Complete the sentences with the correct words from the brackets*		
	(a) ___If___ (Unless, If, Perhaps, Although) you tidy up your room, you won't get candy.	(a)	(a)
	(b) ___Because___ (When, If, Because, Although) I was telling the truth, my mother didn't believe me.	(b)	(b)

No.	Question	(A) Mark the pupil's answer correct or incorrect here	(B) Write the correct answers here
1	Write the missing numbers in the box below		
	(a) 44 + 33 = 77	(a)	(a)
	(b) 86 − 58 = 38	(b)	(b)
	(c) 343 + 215 + 127 = 685	(c)	(c)
	(d) 72 ÷ 9 = 7	(d)	(d)
	(e) 37 × 13 = 3711	(e)	(e)
2	Which two numbers add up to make 0.81? 　(0.24)　0.09　0.90　0.51　(0.57)　0.17	(f)	(f)
3	Circle the one that gives the smallest answer? (a)(39 ÷ 5) (b) 39 ÷ 4 (c) 39 ÷ 3	(g)	(g)

Source: SDI (Service Delivery Indicators) education surveys.

Consistent with the findings from previous work, teacher knowledge and pedagogical skills are also low in this SDI sample. The average teacher in any of the countries cannot correctly answer more than 80 percent of the math and 65 percent of the language knowledge assessment questions. Teacher performance on the three tasks related to pedagogical skills is even lower, with the average teacher in any country not being able to answer more than 40 percent of the questions correctly on any given task. Figure 3.5 shows the average percentage of correct answers by task and the percentage of teachers by formal training for each country. Teacher knowledge and skills are not always correlated with the highest level of formal schooling that teachers have completed (that is, primary or less, secondary, or above secondary), which aligns with the literature on this subject.[15] For instance, whereas Kenya has a high average score for teacher knowledge and a high percentage of teachers with an education above secondary, Nigeria has a relatively low score for teacher knowledge even with a higher percentage of teachers with training beyond secondary school.

66

FIGURE 3.5 Teacher subject and pedagogy knowledge and training in nine African countries, by country

a. Teacher knowledge of subject and pedagogy

b. Teacher training
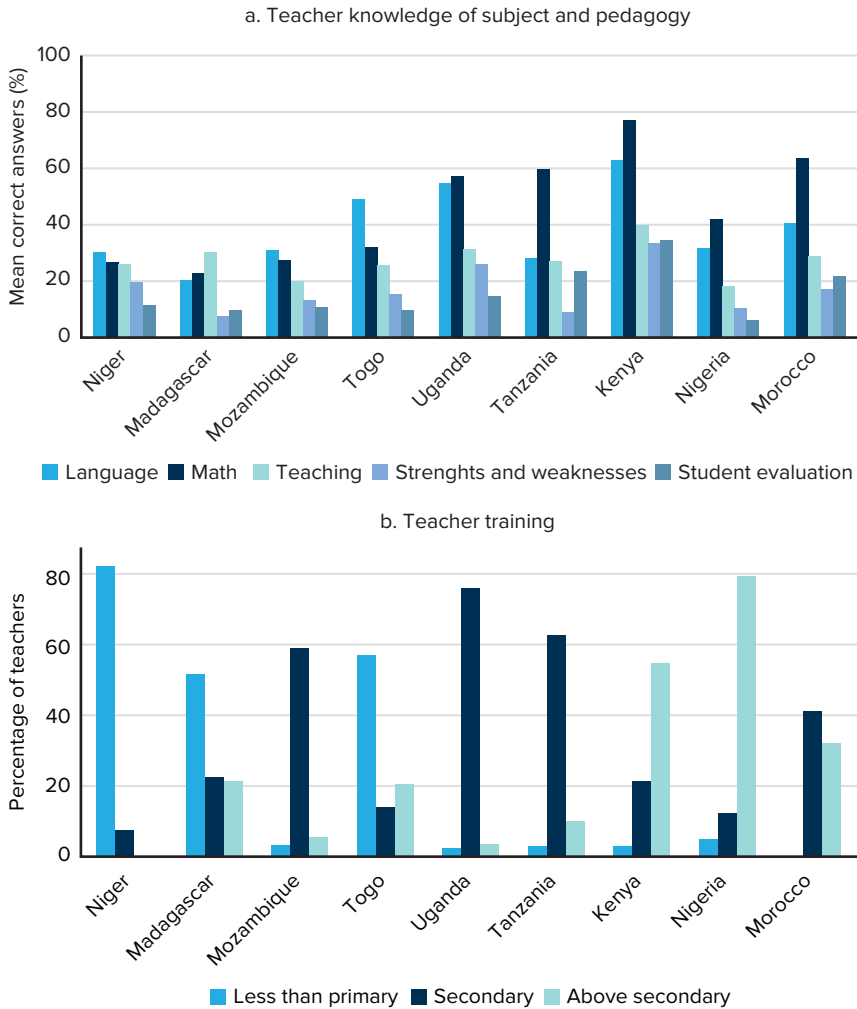
Source: SDI (Service Delivery Indicators) education surveys.
Note: Panel a shows the share of correct answers per task or question on the SDI teacher assessment for public school teachers, sorted by country. Panel b shows the share of teachers by country with different levels of training. Both panels show unweighted means by school. Categories do not always add to 1 because some categories are missing. Country surveys were conducted in the following years: Kenya (2012), Madagascar (2016), Morocco (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Tanzania (2016), Togo (2013), and Uganda (2013).

This type of evidence can be a force for change. An SDI survey in Niger in 2015 found that teacher competencies are weak in comparison with other African countries and that a small fraction of the teachers tested have the minimum level of knowledge to teach French and mathematics at the primary education level. Local researchers conducted the SDI analysis, with the Ministry of Primary Education preparing the report and the National Institute

for Statistics leading data collection and entry. The results paved the way for generating additional data on teaching quality, and the Ministry of Primary Education launched a large-scale national assessment of primary school contractual teachers (who represent 75 percent of the teaching staff). Finally, the evidence from the SDI surveys fed into the World Bank's technical assessment and informed the design of the Learning Improvement for Results in Education (LIRE) project, which has a strong focus on improving the quality of education services through teacher training, coaching, and supervision.

Do schools have essential inputs?

Basic school inputs can encourage children's attendance and may positively affect learning when they support productive teacher-learner interactions (World Bank 2018). A student cannot learn properly if a school lacks minimal infrastructure. Having classrooms in which students can read from the board or where pupils have something to write on are conditions taken for granted in some contexts, but these items are not always available to students in low-resource settings. Factors like student-teacher ratios are also key. Learning is not as effective when classes are too large or extremely diverse in terms of school readiness (Ganimian and Murnane 2016). Similarly, female students may miss school more often or not attend at all if schools do not have a toilet that works and is accessible, clean, private, and separated for boys and girls (Adukia 2017). These infrastructure characteristics also represent the working conditions for teachers—it is extremely difficult for teachers to apply their skills when basic inputs are not in place.

Figure 3.6 visualizes the average percentage of schools with basic inputs, by country. Although most schools across all SDI countries have a functioning blackboard (a board with chalk that is visible to all students), the number of schools with toilets that are gendered and private is often low, and an even lower number of schools have one teacher or more for every 35 students. The percentage of schools that have all of these three inputs fluctuates from 4 percent in Togo to 70 percent in Kenya, but it is below or about 20 percent in six out of the nine countries in this SDI sample. This low percentage shows that there is considerable space for government action to equip every school with the basic inputs needed to improve the quality of education.

Using SDI data, it is possible to characterize how schools with relatively good average learning outcomes differ from those with relatively poor learning outcomes in terms of basic school inputs and teacher characteristics. Although these are not the only factors that influence learning, there is evidence that teacher quality and certain school inputs matter in specific contexts (Araujo et al. 2016; Evans and Popova 2016; Hanushek and Rivkin 2006; World Bank 2018). However, especially in low- and middle-income African countries, evidence is currently

FIGURE 3.6 Basic school inputs in nine African countries, by country and type of input
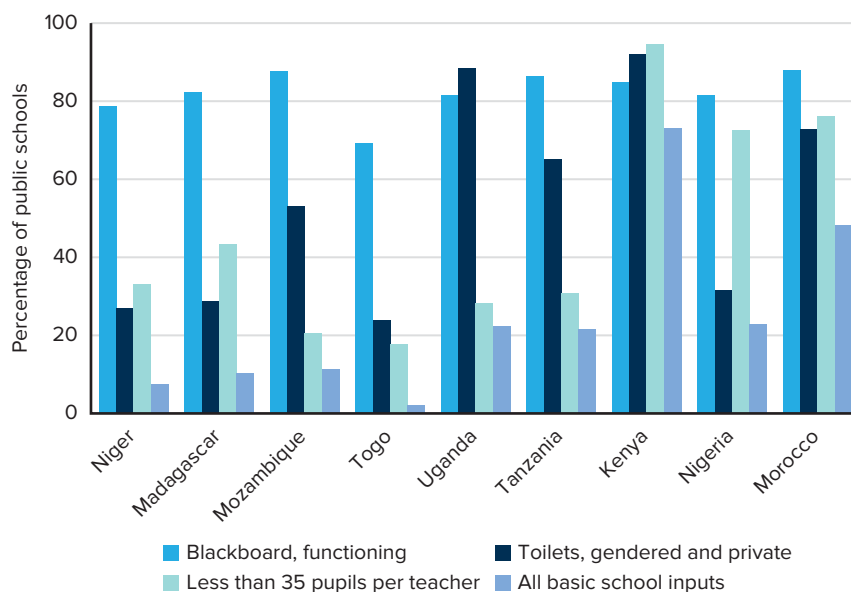
Source: SDI (Service Delivery Indicators) education surveys.
Note: The figure shows the share of public schools by country that have each basic school input and all inputs combined. Country surveys were conducted in the following years: Kenya (2012), Madagascar (2016), Morocco (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Tanzania (2016), Togo (2013), and Uganda (2013).

limited regarding *which* teacher characteristics or school inputs may have the greatest impact on outcomes.

Teacher quality—as reflected in time-on-task, teacher absence, and teacher knowledge—is markedly low in a subset of SDI countries (see Bold et al. 2017, which also uses SDI data). It is plausible that prioritizing improvements in these factors might boost student test scores when systems are starting from low levels. Motivated and well-trained teachers are more likely to be effective when they have additional school inputs (Mbiti et al. 2019). Ultimately, identifying areas of potential action to improve student experience and learning can help to strengthen education systems in low- and middle-income countries in Africa and elsewhere. This direction is important for future analysis.

Learning from what works: What can countries learn from their own success?

Improving education systems in a relatively short period of time is challenging but feasible, and there are many success stories.[16] However, institutional and cultural differences often make it very hard to adapt and scale up lessons learned from

other countries. The large sample of schools included in the SDI surveys makes it possible to analyze those schools that do particularly well within each country so as to learn from these positive outliers. Understanding what the best-performing schools are doing right (and, by implication, potential areas of improvement for poorly performing schools) may help governments to boost learning outcomes.

Understanding high-performing schools

One easy and intuitive way to classify high- and low-performing public schools is to look at the test scores of their students and study two groups of schools in particular: those in the top and bottom 5 percent of the aggregate distribution across all SDI countries. Simply put, this approach yields a group of schools where students are performing well and another group where students are performing poorly. In the highest-performing schools (top 5 percent), the average student masters at least 9 out of 15 math items, whereas in the lowest-performing group (bottom 5 percent) the average student masters none.[17]

The highest-performing and lowest-performing schools are clustered in a few countries. The vast majority of the lowest-performing schools are in just three countries, primarily Nigeria. Meanwhile, the highest-performing schools are clustered in five countries, led by Kenya, Madagascar, Morocco, and Tanzania. Interestingly, Nigeria has a sizable share of both high-performing and low-performing schools, indicating a wide variation in school quality within the country.

Some noteworthy patterns are evident in the characteristics of these high- and low-performing schools. First, high-performing schools are found in both urban and rural areas, but the lowest-performing schools are concentrated in rural areas. The correlation between student test scores and urban-rural locality is positive and significant after controlling for country fixed effects (0.29 of a standard deviation higher test scores for urban schools).[18] However, the distribution shows that high performance is not the exclusive privilege of urban schools. Some rural schools can and do deliver strong results.

Second, there are clear differences regarding school inputs.[19] The highest-performing schools do significantly better on almost every school input. For instance, schools in the highest-performing group are 20 percentage points more likely to have a functioning blackboard, equipment to write on it, and visibility from all seats within the classroom. In terms of pupil-teacher ratio, the highest-performing schools have around 35 students per teacher, whereas the lowest-performing schools have an average of 56 students per teacher (see appendix A, table A.7 for details).

The highest-performing schools not only have more teachers per student than their counterparts in the lowest-performing schools, but their teachers are also better prepared. Teachers in the highest-performing schools score 1.3–1.4 standard deviations higher on numeracy and literacy knowledge, respectively, and 0.65 of a standard deviation higher on pedagogy knowledge. The top group

of schools also retains teachers who are, on average, older by about four years and more likely to be female by 40 percentage points.

What may be driving differences in school performance?

Comparing high- and low-performing schools, tables A.7 and A.8 in appendix A highlight the specific teacher characteristics and school inputs that appear to distinguish the highest-performing schools most clearly from schools at the bottom of the distribution. Among inputs, lower pupil-teacher ratios in top-ranking schools are a salient distinguishing factor, with high-performing schools having 20 fewer students per teacher on average, as published evidence would lead one to predict (Chetty et al. 2011).[20] But simpler inputs, such as the presence of functioning blackboards and clean, private, gender-separated toilets also appear to play a role. The analysis also shows substantial differences in teacher test scores between high- and low-performing schools, an intuitive result. But striking differences also emerge concerning the gender and age composition of the teaching workforce. The following section explores these key differentiating factors in detail. Doing so suggests entry points for policy to reduce performance gaps between these groups of schools.

High- and low-performing schools: How can countries narrow the gaps?

The comparison of highest- and lowest-performing schools yields findings that can inform policy to boost performance among lagging schools. Some basic school inputs are particularly scarce in low-performing schools, which suggests that, once in-person learning resumes, governments have room to make meaningful improvements with simple solutions that matter for learning outcomes.

SDI results contain information helpful for targeting interventions to improve school inputs. For instance, it is possible to look at how many schools by country or by urban-rural setting have some essential features or "minimum inputs" to operate efficiently. Minimum inputs are defined here as having functioning blackboards, private and gendered toilets, and a pupil-teacher ratio below 35. Across SDI countries, only 20 percent of rural schools possess these minimum school inputs, and 35 percent of urban schools meet this basic standard. Thus, in some settings, targeting rural schools early for additional inputs could deliver results. In some countries, SDI data on school inputs may serve as a more generalized wake-up call to decision-makers. In Togo and Niger, for example, only 2 percent and 8 percent of all schools, respectively, have the minimum school inputs as measured by SDI surveys.

SDI results point to tangible actions for in-person learning that policy makers can take to address school inputs. Yet caution is warranted, because

the evidence in this area suggests that more or better resources improve student achievement only if they translate into real changes in children's daily experiences at school (Ganimian and Murnane 2016). On average, other factors remaining constant, adding one teacher per school and ensuring that every school has a functioning blackboard and private and gendered toilets is associated with an increase of 0.24 of a standard deviation in students' average test scores (see table A.9 in appendix A).[21] Although this evidence does not give insights into the causal pathways for reform, it still points to feasible solutions with clear benefits for the order of magnitude of the impact generated by the most successful interventions to improve test scores at scale (see Glewwe and Muralidharan 2016; McEwan 2015). Some of these improvements, such as the

---

BOX 3.2   Toilets and handwashing facilities in schools: A key concern during COVID-19

In many respects, the COVID-19 (coronavirus) pandemic has reshaped the way people think about education. Millions of schools have shut down and countries are continuously reassessing whether it is safe to reopen them. An essential condition to allow schools to reopen is ensuring a safe return for students and teachers, which means being able to maintain physical and social distancing as well as implementing public health measures like frequent handwashing. The availability of basic sanitation infrastructure such as clean toilets is essential to prevent or reduce the spread of COVID-19 and other diseases. Some of these features might be difficult to attain in low- and middle-income contexts—such as those in the Service Delivery Indicators (SDI) sample—where basic sanitation infrastructure is already lacking.

SDI results provide information on some of the sanitation inputs that have gained importance in the current context. Although some SDIs date back several years, more recent ones—including the 2016 surveys for Madagascar Morocco, and Tanzania—may give policy makers a good indication of what investments are needed to facilitate a safe reopening. In particular, the data allow analysts to look into (1) the availability of toilets for students, (2) the availability of a clean toilet, and (3) the availability of a handwashing facility with soap and water near the toilets.

SDI data show that, generally, investments in sanitation infrastructure are needed more urgently in rural settings, although schools located in urban areas also need support. Although the presence of at least a toilet seems to be the norm in most schools, there are still gaps, and not every school has one. These shortfalls may be largest in rural schools in Madagascar and Morocco, where 68 and 76 percent of schools have at least one toilet, respectively, as figure B3.2.1 shows. However, there is more room for improvement with regard to cleanliness. In the three countries with 2016 SDI surveys, only 65 percent of schools have a toilet assessed as clean.

Of increased importance in the setting of COVID-19 is the presence of handwashing stations near these toilets. As shown in figure B3.2.1, this feature was not prioritized in the past. Overall, only 48 percent of schools in the three countries have a handwashing station in proximity to their sanitary facilities. This means that more than 50 percent of students are not able to wash their hands after using the toilet. This inability represents a serious health risk, especially in the current context.

Although SDI surveys were not designed with these issues in mind, the richness of the data generated makes it possible to shed some light on this and other characteristics that can help policy makers to make appropriate decisions. As the SDI team continues to revise and enhance its instruments for upcoming surveys, more questions on this front will likely be needed. Chapter 4 explores some future survey plans in greater detail.

(Continued)

BOX 3.2   Toilets and handwashing facilities in schools: A key concern during COVID-19 *(Continued)*

FIGURE B3.2.1   Toilets and handwashing facilities in public schools in three African countries, by country, 2016
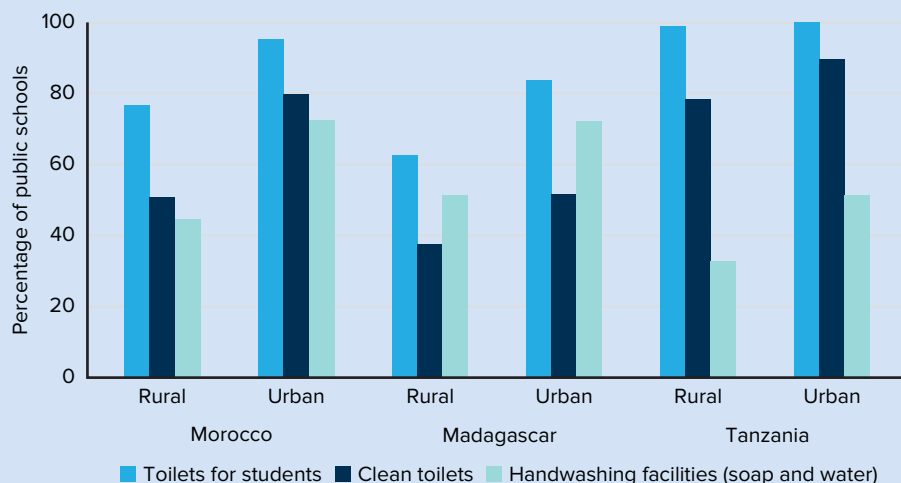
Source: SDI (Service Delivery Indicators) education surveys.
Notes: The figure shows the percentage of public schools in the three countries with the most recent SDI education surveys that have toilets and handwashing facilities.

presence of handwashing facilities in schools, have acquired renewed importance in the wake of COVID-19 (box 3.2).

Leveraging teacher traits to improve results

The comparison exercise between the highest-performing and lowest-performing schools also shows stark differences in teacher characteristics. Numerous studies suggest that teachers' knowledge matters for student learning, particularly when the knowledge concerns the specific content that instructors are responsible to teach and pedagogical tasks that they should perform on a regular basis (see, for instance, Hill, Rowan, and Ball 2005; Metzler and Woessmann 2012). Other characteristics that are easier to observe, such as formal education and accreditation, have been shown to have no link to better teacher performance (Bau and Das 2020; Hanushek and Rivkin 2006).

Teacher characteristics vary substantially across SDI countries. On average across all countries, 13.6 percent of teachers have a primary education or less, 50.5 percent have completed secondary schooling, and 35.4 percent have

completed a postsecondary degree.[22] Niger has the highest percentage of teachers with a primary education or less (86.2 percent), whereas Nigeria has the most educated teachers, with the vast majority having a postsecondary degree (80.3 percent). Large differences in pedagogy, numeracy, and literacy between the highest- and lowest-performing schools are evident. Surprisingly, teacher effort does not seem to be a critical factor in learning outcomes, although this may be because the proxy measure for effort—teacher absence—captures only certain dimensions of effort. However, teachers' knowledge is a strong predictor of student learning. Consistent with the literature, teacher quality is not strongly linked to teachers' formal education and accreditations in the SDI data.

Other observable characteristics such as gender and age, likely a proxy for experience, are strong predictors of teachers' effectiveness in SDI countries. Students in schools with more female teachers perform better across the SDI sample of public schools. This positive association is driven in great part by schools located in rural areas and schools with older female teachers. One potential explanation for

FIGURE 3.7 Average association between teachers' gender and age and students' learning performance in nine African countries

a. Student performance on math tests and teachers' gender

b. Student performance on math tests and teachers' age

Percentage of teachers who are female

Age of teachers

● 25% of teachers are female
■ 75% of teachers are female
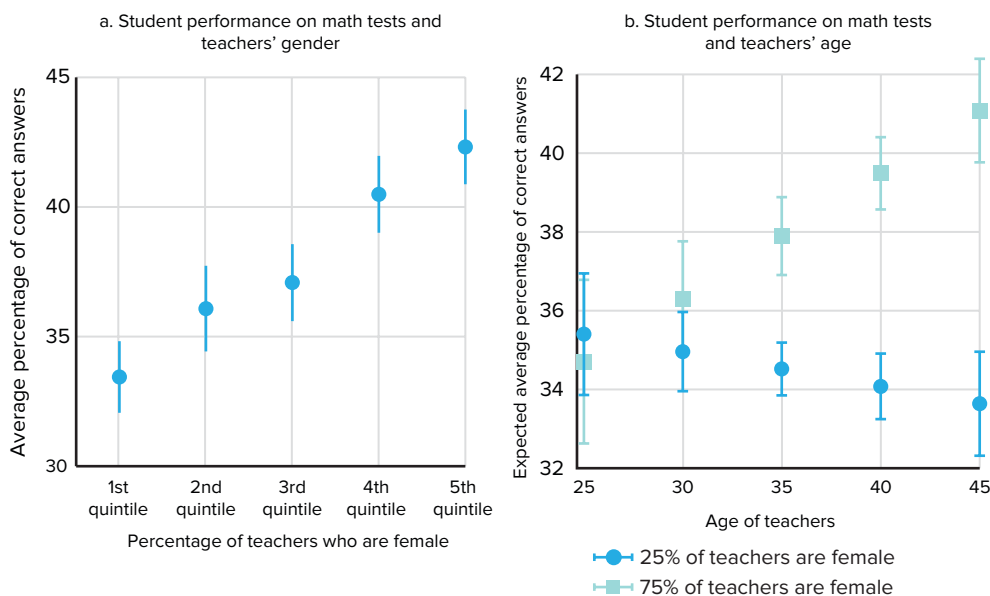
Source: SDI (Service Delivery Indicators) education surveys.
Note: Panel a shows the average share of correct answers on the SDI math assessment for students in public schools, grouped by within-country quintiles according to the share of female teachers in schools. Panel b shows the linear prediction for the average share of correct answers on the SDI math assessment for students in rural public schools, grouped by the average age of teachers for schools with a share of female teachers of 25% and 75%. The estimates are from a regression controlling for teachers' formal training and country fixed effects. Each linear prediction shows 95% confidence intervals with robust standard errors. Country surveys were conducted in the following years: Kenya (2012), Madagascar (2016), Morocco (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Tanzania (2016), Togo (2013), and Uganda (2013).

this association is that schools with a higher share of female teachers display lower levels of teacher absence. In fact, teacher absenteeism is lower for female teachers, with the gender gap widening as their age increases across all schools. In turn, teachers' formal education contributes very little to explain differences between top- and bottom-performing schools. Results are shown in figure 3.7.

Across the SDI sample, older female teachers perform better than their counterparts on key measures that correlate positively with higher pupil test scores, including teacher pedagogy, numeracy, and literacy scores as well as lower teacher absence rates. These correlations are stronger in rural settings.[23] In some countries, teaching represented for a long time one of the few career paths deemed culturally appropriate for women (Goldin 2015). This might explain why a relatively large number of qualified, motivated women enters the teaching profession, and this situation is reflected in generational trends. Although the analysis to date cannot rule out sorting of some types of teachers into selected schools, ideally, governments might want to attract teachers whose qualifications and motivation resemble those of the older female teachers in the SDI sample.

Learning from private schools

The earlier sections of this chapter focused on public education, which is more easily comparable across the SDI countries in the sample. Instead, private schools and the rules governing their operation tend to differ widely from one country to another, even within the SDI sample. In Uganda, for example, some private schools receive government funding and, as a result, are subject to some rules and regulations; this is not the case in all SDI countries. Substantial variation in the private share of the education market by country and urban-rural locality also makes direct comparisons hard to justify and interpret. For instance, in Niger, rural private schools were not included in the sample frame because of their scarcity. For the same reason, in Morocco, virtually no rural private schools were included in the final sample. Thus comparing private schools in either of these two countries with those in other SDI countries would implicitly limit one side to only urban schools.

Notwithstanding these challenges, studying the private education sector can yield valuable lessons for improving student learning outcomes. To explore differences between public and private schools in a meaningful way, the analysis that follows focuses on a subsample of countries where both private and public school sectors are large enough to generate reliable results and where both urban and rural schools of both types can be included. These countries are Kenya, Madagascar, Nigeria, and Togo.

Across these four countries, public schools represent about 60 percent to 78 percent of each country sample, and the remaining schools are classified as private. The composition of the private sector, however, varies depending on the context. For instance, in Kenya, 15 percent of schools are private for-profit, 7 percent are private not-for-profit, and the remaining 78 percent are public.

In Madagascar, 64 percent of schools are public, 12 percent are private with religious affiliation, and 23 percent are private without religious affiliation. In Nigeria, 3 percent of schools in the sample are private community schools, 7 percent are private not-for-profit, and 30 percent are private for-profit; the remaining 60 percent are public. Finally, in Togo, 13 percent of the sample are private schools without religious affiliation, 14 percent are private with religious affiliation, and the remaining 73 percent are public. Despite these differences, some general trends emerge when comparing private and public schools.

Results, shown in figure 3.8, indicate that, in Kenya, Madagascar, Nigeria, and Togo, private schools do better than public schools, on average, as captured by better performance on the math test by children attending private schools. The exercise previously described of grouping highest- and lowest-performing schools was repeated for this subset of countries, but now including private schools. Some suggestive patterns emerge from this analysis. First, there are different types of high-performing schools among all of the possible combinations of urban-rural and private-public groups of schools. Urban private schools make up the largest single share of the highest-performing schools (40

FIGURE 3.8   Differences in math test scores between public and private schools in four African countries, by country
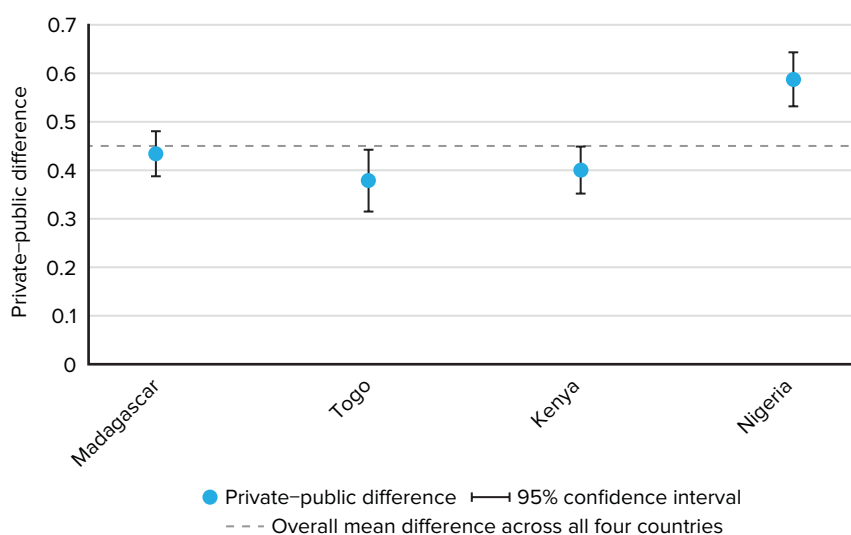
Source: SDI (Service Delivery Indicators) school surveys.
Note: This figure shows the average difference between the SDI math assessment results of students in private and public schools. Values are obtained from the coefficient of regressing math test scores on a private school dummy. Test scores are equated across the whole set of countries using an IRT 2-PL model and MLE estimates. Only countries with a relatively balanced public and private sector are included. The 95% confidence interval is shown with robust standard errors. Country surveys were conducted in the following years: Kenya (2012), Madagascar (2016), Nigeria (2013), and Togo (2013). IRT 2-PL = item response theory two-parameter logistic. MLE = maximum likelihood estimation.

TABLE 3.2 Comparison of highest- and lowest-performing public and private schools in four African countries, by rural-urban location

	Lowest performers				Highest performers			
Variable	Rural pubic	Rural private	Urban public	Urban private	Rural public	Rural private	Urban public	Urban private
Test scores	−2.10	−1.59	−1.59	−1.11	1.86	1.62	1.43	1.56
Principal component analysis, first component								
Teacher characteristics	−0.54	−0.72	0.15	−1.31	−0.12	0.50	−0.95	0.00
School inputs	−1.25	−0.89	−1.32		0.05	0.49	−0.09	0.66
Number	141	19	13	1	48	48	8	69

Source: SDI (Service Delivery Indicators) education surveys.
Note: This table compares schools in the top and bottom 10% of the distribution for Kenya, Madagascar, Nigeria, and Togo and includes both public and private schools. The first component of the principal component analysis decomposition contains the most variation of the variables included. Country surveys were conducted in the following years: Kenya (2012), Madagascar (2016), Nigeria (2013), and Togo (2013).

percent). There are almost no urban private schools and very few rural private schools among the lowest-performing schools.

Within the lowest-performing group of schools, the few private schools present are doing better than public schools (table 3.2). However, among the highest-performing group, public schools are doing as well as or better than private schools. Although small sample sizes for certain groups make some of the comparisons merely indicative, differences in school inputs and teacher characteristics are unlikely to explain the differences in learning between private and public schools.

One possible reason why private schools do better overall in terms of learning is that they might have a higher share of wealthier students, who can potentially take advantage of better conditions for learning at home. In this case, the difference in learning outcomes between public and private schools should disappear once the analysis takes into account the differences in socioeconomic status among students. Unfortunately, measures of household socioeconomic status were not collected for the SDI countries in this sample. Although the most recent SDI surveys are now collecting this information, previous SDI surveys have geo-references that may be merged with geospatial data to address some of these questions (see chapter 4 for details). In this case, nighttime lights can be used as a proxy for the socioeconomic status of school location. Expectedly, students in more affluent locations do better, but the difference between private and public schools stays constant throughout the spectrum of welfare (figure 3.9), suggesting that the difference between public and private schools in the sample cannot be explained entirely by the socioeconomic composition of the student body. Nonetheless, we cannot discard the presence of within-area sorting of students, and more accurate measures of students' household socioeconomic status would help to gain clearer insights in the future.

FIGURE 3.9 Prediction of student outcomes over the range of a proxy for socioeconomic status in nine African countries, public schools versus private schools
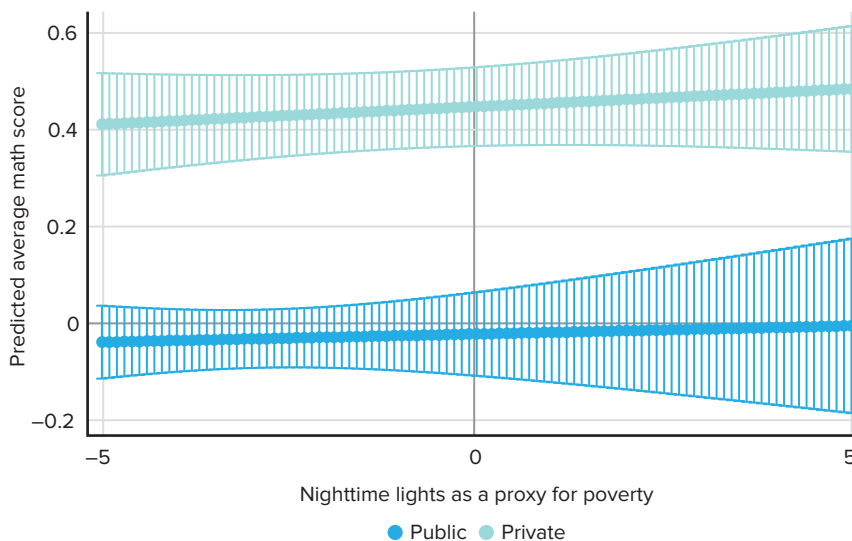
Source: SDI (Service Delivery Indicators) school surveys.
Note: The figure shows the linear prediction for the SDI math test scores across the distribution of nighttime lights (a proxy for poverty), controlling for urban-rural locality, a wide array of teacher characteristics (age, gender, formal training), and country fixed effects. Each linear prediction shows 95% confidence intervals with robust standard errors. Country surveys were conducted in the following years: Kenya (2012), Madagascar (2016), Morocco (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Tanzania (2016), Togo (2013), and Uganda (2013).

There are many potential drivers behind the difference in learning between public and private schools. For instance, evidence from the OECD countries finds that private schools exhibit, on average, better management systems that tend to be positively correlated with learning (OECD 2012). Better management could explain some of the public-private gap in learning outcomes observed in this sample. Unfortunately, as with household socioeconomic status, SDI surveys historically have not collected this information. Chapter 5 describes some recent SDI efforts to capture these two dimensions. The new approaches hold promise for clarifying this and other policy-relevant correlations with a view to improving student learning outcomes in low- and middle-income contexts.

Conclusions: What will it take to improve service delivery in education?

A comprehensive review of all of the policies that are necessary to improve the quality of education, during and after COVID-19, is beyond the scope of this book.

Various World Bank Group flagship reports have surveyed the literature of causal evidence, analyzed the complex educational landscape amid the pandemic, and suggested a way to move forward and protect the learning and schooling gains and build back better education systems (World Bank 2018, 2019, 2020). That said, SDI evidence offers unique insights into margins that could significantly improve students' school experience and learning, particularly in the African context. In combination with previous evidence in the field, this analysis can be translated into practical guidance for policy makers who need to prioritize investments to "build back better" in education, minimizing learning losses now, while securing the human capital on which countries' long-term prosperity depends.

The analysis presented in this chapter supports a number of broad directions for action. Each country can and should shape the principles to its own specific context, recognizing what is already working and what is not.

- *Identify top-performing schools, analyze how they succeed, and share their solutions.* Despite generally low average student test scores in SDI countries to date, wide within-country variance exists, and each country has high-performing schools. Adapting lessons from other countries is often challenging because of cultural and institutional differences. Extrapolating lessons from top-performing public schools might constitute an easier first step for countries wanting to improve their learning outcomes. By providing comprehensive information on frontline aspects of education systems, SDI surveys can help governments to deploy investments so that larger numbers of young people can reach their full potential.

- *To improve inputs, start with the basics—but do not stop there.* Consistent with the findings from previous impact evaluations (see Ganimian and Murnane 2016 for a literature review), SDI results indicate that specific school inputs can potentially improve students' daily experience in school, such as blackboards and functioning toilets. These investments, although taken for granted in some settings, can be more effective at enhancing the experience of students and teachers than expensive infrastructure. For education systems aiming to build back better, these basic inputs would be a good place to start. In most cases, however, such inputs alone will not be sufficient to deliver the educational quality that citizens increasingly expect. These basic measures must be coupled with other policies that align incentives, increase accountability, and are framed within comprehensive programs.

- *Hire more and better teachers.* Schools with better-performing students in the SDI sample have more teachers per student and also more knowledgeable teachers. Accordingly, policy makers looking to make the most effective and efficient use of resources may also be looking to hire more and better teachers. The existing literature suggests that, in

low- and middle-income countries, lowering the pupil-teacher ratio might be effective only under certain circumstances (Chetty et al. 2011; Duflo, Hanna, and Ryan 2012; Ganimian and Murnane 2016). However, if incentives are aligned and coupled with the necessary complementary interventions,[24] class size emerges as a strong correlate of differences in performance across schools in the SDI sample.

- *Recruit teachers on the basis of talent, not solely titles.* As governments try to attract motivated professionals into the teaching career, they should prioritize recruitment systems that privilege measures of performance rather than static attributes such as education and professional qualifications.

- *Support efforts that encourage teachers to be learners.* Perhaps one of the most challenging tasks for education authorities is to implement professional learning systems that support teachers' continuous growth in content knowledge and pedagogical skills, two strong predictors of students' learning.

- *Private school solutions can spark public sector innovation.* Evidence presented in this chapter suggests that private schools in SDI countries tend to have higher average levels of learning than public schools. As a result, much can be learned from studying the private educational sector. Such analyses may yield lessons that can be applied in public systems or that stimulate public institutions to innovate and create their own solutions. For example, although private schools might be able to attract better school principals by paying higher wages, recent evidence for Chile shows that public schools, despite relatively rigid wage structures, can attract better principals by improving selection criteria, with more qualified principals contributing to better learning outcomes (Muñoz and Prem 2020).

The COVID-19 pandemic and its ensuing economic crisis will have important medium- and long-term effects on student learning outcomes. Children who were forced temporarily or permanently out of school will need to be given the opportunity to catch up. When timely and usable administrative data are available, countries can and have been taking steps to identify and target interventions directed at students who are at high risk of dropping out or lagging behind on learning. For instance, teaching at the right level or providing remedial tutoring is most effective if targeted at the most disadvantaged and at-risk student groups. However, these data are rarely available in more challenging settings. In the years ahead, new SDI surveys will help to bridge this gap and support policy makers in targeting interventions to the groups most in need, particularly when other information to do so is not readily available.

As countries grapple with the consequences of COVID-19, education systems have the opportunity to emerge better, stronger, and more prepared to support students under emergency conditions and in the "new normal" the

crisis will leave behind. To achieve this improvement will require strategic leadership informed by evidence and analysis of the main drivers of learning. Consolidating resilient education systems will require decision-makers to make the most of constrained financial resources during an economic downturn whose impacts in many settings are likely to be prolonged. SDI surveys can contribute through the evidence and lessons already produced. As the effort evolves, it will also support countries to continue transforming their education and other social service systems, applying new tools and methods to spark progress for students, teachers, families, and societies. The next chapter looks in detail at some of the measurement innovations that can drive this work.

Notes

1 | For simplicity, this book presents unweighted results. However, all messages extracted from the data are robust to using weights.

2 | Morocco is not included in this estimation because it is not part of Sub-Saharan Africa. This sample focuses on SDI surveys that were completed between 2012 and 2018. The 2010 Senegal and Tanzania pilot surveys are excluded because they were conducted at a smaller scale and are not fully comparable. Similar exercises such as the Systems Approach for Better Education Results (SABER)–Service Delivery surveys are not considered for this book, because they have not been fully harmonized with SDI data. Recently collected data that have not yet been validated will be included in future reports. When panel data are available for a given country, the latest year is used. The share of schoolchildren is estimated on the basis of each country's latest available statistics from the World Bank (https://databank.worldbank.org/).

3 | The estimates of GDP per capita (based on purchasing power parity in current international dollars) come from World Bank Open Data, and the year of the survey is used for each country's estimate. See https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD

4 | See appendix D for details on the methodological groundwork for the SDI teacher and student assessments.

5 | The sample of SDI countries in Bold et al. (2017) differs from the one used in this book. In particular, that paper uses data from the SDI pilots in Senegal and Tanzania in 2010 and from Tanzania in 2014, whereas this book excludes those data in favor of newer data for Tanzania (2016) and adds data for Madagascar (2016), Morocco (2016), and Niger (2015). For more details on the selection of surveys for this book, see the discussion on sample, methods, and framework.

6 | The difference is 0.07 of a standard deviation in a regression of math test scores, controlling for student age and country fixed effects.

7 | For the language component of the student assessment, a subsample of students in some countries was tested in a language other than the language of instruction, typically the vernacular. Although this approach provides important insights for within-country analysis and policy recommendations, it makes language scores less comparable across countries. For that reason, the remainder of this subsection focuses only on students' mathematics test scores when presenting cross-country comparisons. For further details, see box 3.1.

8 | In order to make test scores comparable across countries and time and to put them on the same scale, test scores are computed using psychometric linking methods from item response theory and then standardized to have mean = 0 and standard deviation = 1. Maximum likelihood estimates are used throughout this book. For details on these methods, see Bau, Das, and Yi Chang (2021); Das and Zajonc (2010).

9 | In contrast, Harmonized Learning Outcomes rankings and differences in test scores for this group of countries show Kenya as having the highest performance and Niger as having the lowest performance (and a difference of 1.5 standard deviations).

10 | A simple ANOVA decomposition shows that, whereas country fixed effects explain up to 63.5 percent of the variation, the first principal component obtained through principal component analysis of school inputs and teacher characteristics accounts for 6.6 percent and 11.5 percent of the explained variation in the SDI sample.

11 | The characteristics that this chapter focuses on were selected mostly through a double-lasso regression on student test scores. A very similar combination of variables was prioritized by the first principal component obtained through principal component analysis as explaining the most variation within four categories of variables corresponding to the World Development Report 2018 framework.

12 | See the discussion in chapter 1.

13 | These averages are in line with the 23 percent and 44 percent reported in Bold et al. (2017). However, new countries provide different pictures. For instance, teacher absence from school and the classroom in Morocco seems to be low, at 3 percent and 4 percent, respectively, but only after excusing those who were initially reported as absent because they were working on a different shift. This correction substantially affects only the estimates for Morocco, which otherwise would have a teacher absence rate of 23 percent from both school and classroom. The earlier paper uses data from the SDI pilots in Senegal and Tanzania in 2010 and from Tanzania in 2014; this book excludes those data in favor of newer data for Tanzania (2016) and adds data from Niger (2015), Madagascar (2016), and Morocco (2016). This book also focuses on public schools (in the current subsection), whereas the results of Bold et al. (2017) include both public and private schools. Furthermore, the use of different weights for some countries and the reclassification of some teachers reported to be absent because the visit did not take place during their shift might create a small difference in the estimates. More details on how SDI countries were selected for this book can be found in the section on samples, methods, and framework.

14 | It is difficult to gauge the reliability of this information and the degree to which it appropriately explains the bulk of teacher absences reported in the Morocco survey. Future SDI surveys could perform a follow-up investigation in a small sample of facilities to study the share of excused absences that are legitimate.

15 | A regression of assessment scores and education-level groups of teachers finds significant and positive, albeit very small, associations (2–4-percentage-point differences) between groups of teachers by level of education and teachers' knowledge and pedagogical skills after controlling for country fixed effects, urban-rural locality, and teachers' age and gender.

16 | A substantial number of such stories is documented in *World Development Report 2018: Learning to Realize Education's Promise* (World Bank 2018).

17 | "Mastering" refers to having more than 70 percent of correct answers across all students in the school group.

18 | Obtained through a simple multivariate regression at the school level of mean student test scores on an urban-rural indicator controlling for country fixed effects and with robust standard errors.

19 | These differences are obtained by doing a simple *t*-test of means between the top and bottom 5 percent of public schools.

20 | This finding does not necessarily contradict the results of Duflo, Dupas, and Kremer (2015), who find that lowering class size by adding more centrally hired civil service teachers does not improve student learning outcomes. They attribute this null effect to existing teachers' reducing their effort in response to the new hires and helping their relatives to get hired into a significant portion of the new teaching slots.

21 | Results from a multivariable linear regression of student test scores on school inputs controlling for country fixed effects and proxies for poverty and accessibility to school.

22 | The remaining 0.5 percent corresponds to the category "other."

23 | These results come from regressions at the teacher level controlling for school locality, teacher formal training, and country fixed effects.

24 | For instance, in Tanzania a joint intervention providing teacher incentives coupled with school grants has had significantly larger effects on student test scores than providing incentives or grants separately (Mbiti et al. 2019).

References

Adukia, A. 2017. "Sanitation and Education." *American Economic Journal: Applied Economics* 9 (2): 23–59.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. *Standards for Educational and Psychological Testing*. Washington, DC: AERA.

Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo, and N. Schady. 2016. "Teacher Quality and Learning Outcomes in Kindergarten." *Quarterly Journal of Economics* 131 (3): 1415–53.

Azevedo, J. P., and D. Goldemberg. 2020. "Learning for All: Within-Country Learning Inequality." *Education for Global Development* (blog), November 12, 2020. https://blogs.worldbank.org/education/learning-all-within-country-learning-inequality.

Bashir, S., M. Lockheed, E. Ninan, and J. P. Tan. 2018. *Facing Forward: Schooling for Learning in Africa*. Washington, DC: World Bank.

Bau, N., and J. Das. 2020. "Teacher Value Added in a Low-Income Country." *American Economic Journal: Economic Policy* 12 (1): 62–96.

Bau, N., J. Das, and A. Yi Chang. 2021. "Learning Trajectories through Primary Schools: New Evidence from Pakistan." *International Journal of Educational Development* 84 (July): 102430.

Bold, T., D. Filmer, G. Martin, E. Molina, B. Stacy, C. Rockmore, J. Svensson, and W. Wane. 2017. "Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa." *Journal of Economic Perspectives* 31 (4): 185–204.

Bold, T., D. Filmer, E. Molina, and J. Svensson. 2019. "The Lost Human Capital: Teacher Knowledge and Student Achievement in Africa." Policy Research Working Paper 8849, World Bank, Washington, DC.

Cameron, S. V., and J. J. Heckman. 2001. "The Dynamics of Educational Attainment for Black, Hispanic, and White Males." *Journal of Political Economy* 109 (3): 455–99.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126 (4): 1593–660.

Collier, V., and W. P. Thomas. 2017. "Validating the Power of Bilingual Schooling: Thirty-Two Years of Large-Scale, Longitudinal Research." *Annual Review of Applied Linguistics* 37 (September): 203–17.

Cruz-Aguayo, Y., P. Ibarrarán, and N. Schady. 2017. "Do Tests Applied to Teachers Predict Their Effectiveness?" *Economics Letters* 159 (October): 108–11.

Cunha, F., and J. Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47.

Das, J., and T. Zajonc. 2010. "India Shining and Bharat Drowning: Comparing Two Indian States to the Worldwide Distribution in Mathematics Achievement." *Journal of Development Economics* 92 (2): 175–87.

Duflo, E., P. Dupas, and M. Kremer. 2015. "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." *Journal of Public Economics* 123 (March): 92–110.

Duflo, E., R. Hanna, and S. P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4): 1241–78.

Educational Testing Service. 2014. "2014 ETS Standards for Quality and Fairness." Educational Testing Service, Princeton, NJ.

Evans, D., and A. Popova. 2016. "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews." *World Bank Research Observer* 31 (2): 242–70.

Filmer, D., E. Molina, and W. Wane. 2020. "Identifying Effective Teachers: Lessons from Four Classroom Observation Tools." World Bank, Washington, DC.

Ganimian, A. J., and R. J. Murnane. 2016. "Improving Education in Developing Countries: Lessons from Rigorous Impact Evaluations." *Review of Educational Research* 86 (3): 719–55.

Glewwe, P., Q. Chen, and B. Katare. 2012. "What Determines Learning among Kinh and Ethnic Minority Students in Vietnam: An Analysis of Round 2 Young Lives Data." Working Paper 80, Young Lives, Oxford, UK.

Glewwe, P., and K. Muralidharan. 2016. "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications." In *Handbook of the Economics of Education*, vol. 5, edited by E. A. Hanushek, S. J. Machin, and L. Woessmann, 653–743. Amsterdam: Elsevier.

Goldin, C. 2015. "A Pollution Theory of Discrimination: Male and Female Differences in Occupations and Earnings." In *Human Capital in History: The American Record,* 313–48. Chicago, IL: University of Chicago Press.

Grisay, A., J. H. A. L de Jong, E. Gebhardt, A. Berezner, and B. Halleux-Monseur. 2007. "Translation Equivalence across PISA Countries." *Journal of Applied Measurement* 8 (3): 249–66.

Hanson, M. 2021. "K–12 School Enrollment & Student Population Statistics." Educationdata.org, July 10, 2021. https://educationdata.org/k12-enrollment-statistics#.

Hanushek, E., and S. Rivkin. 2006. "Teacher Quality." In *Handbook of the Economics of Education*, vol. 2, edited by E. A. Hanushek and F. Welch. Amsterdam: Elsevier.

Hanushek, E. A., and S. G. Rivkin. 2012. "The Distribution of Teacher Quality and Implications for Policy." *Annual Review of Economics* 4 (1): 131–57.

Heckman, J. J., J. Stixrud, and S. Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics* 24 (3): 411–82.

Hill, H. C., B. Rowan, and D. L. Ball. 2005. "Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement." *American Educational Research Journal* 42 (2): 371–406. https://doi.org/10.3102/00028312042002371.

International Commission on Financing Global Education Opportunity. 2016. "The Learning Generation: Investing in Education for a Changing World." International Commission on Financing Global Education Opportunity, New York.

Liu, J., S. Loeb, and Y. Shi. 2020. "More Than Shortages: The Unequal Distribution of Substitute Teaching." *Education Finance and Policy* (blog), September 1, 2020. https://direct.mit.edu/edfp/article/doi/10.1162/edfp_a_00329/97132/More-Than-Shortages-The-Unequal-Distribution-of.

Mbiti, Isaac M. 2016. "The Need for Accountability in Education in Developing Countries." *Journal of Economic Perspectives* 30 (3): 109–32. doi: 10.1257/jep.30.3.109.

Mbiti, I., K. Muralidharan, M. Romero, Y. Schipper, C. Manda, and R. Rajani. 2019. "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania." *Quarterly Journal of Economics* 134 (3): 1627–73. https://doi.org/10.1093/qje/qjz010.

McEwan, Patrick J. 2015. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." *Review of Educational Research* 85 (3): 353–94.

Metzler, J., and L. Woessmann. 2012. "The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation." *Journal of Development Economics* 99 (2): 486–96.

Mullis, I. V. S., M. O. Martin, P. Foy, and M. Hooper. 2017. "PIRLS 2016 International Results in Reading." TIMSS & PIRLS International Study Center, Boston College, Boston, MA. http://timssandpirls.bc .edu/pirls2016/international-results/.

Muñoz, P., and M. Prem. 2020. "Managers' Productivity and Labor Market: Evidence from School Principals." Working Paper 40, Red Investigadores de Economía.

Muralidharan, K., J. Das, A. Holla, and A. Mohpal. 2016. "The Fiscal Cost of Weak Governance: Evidence from Teacher Absence in India." Policy Research Working Paper 7579, World Bank, Washington, DC.

OECD (Organisation for Economic Co-operation and Development). 2004. *Learning for Tomorrow's World: First Results from PISA 2003*. Paris: OECD. https://doi.org/10.1787/19963777.

OECD (Organisation for Economic Co-operation and Development). 2012. "Management and Funding." In *Public and Private Schools: How Management and Funding Relate to Their Socio-economic Profile*. Paris: OECD. https://doi.org/10.1787/9789264175006-5-en.

OECD (Organisation for Economic Co-operation and Development). 2013. *Resources, Policies, and Practices*. Vol. IV of *PISA 2012 Results: What Makes Schools Successful*. Paris: OECD. https://doi .org/10.1787/19963777.

OECD (Organisation for Economic Co-operation and Development). 2018. "PISA for Development: Results in Focus." *PISA in Focus* 91 (December).

Ramachandran, R. 2012. "Language Use in Education and Primary Schooling Attainment: Evidence from a Natural Experiment in Ethiopia." IEB Working Paper 2012/34, Institut d'Economia de Barcelona, Barcelona.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73 (2): 417–58. https://doi.org/10.1111/j.1468-0262.2005.00584.x.

UN (United Nations). 2011. *Millennium Development Goals Report 2011*. New York: United Nations, June 2011. https://www.refworld.org/docid/4e42118b2.html.

UN (United Nations). 2020. "Policy Brief: Education during COVID-19 and Beyond." UN, New York, August 2020. https://unsdg.un.org/resources/policy-brief-education-during-covid-19-and-beyond.

UN (United Nations) General Assembly. 2015. "Transforming Our World: The 2030 Agenda for Sustainable Development." A/RES/70/1, October 21, 2015. UN General Assembly, New York. https://www .refworld.org/docid/57b6e3e44.html.

UNESCO (United Nations Educational, Scientific and Cultural Organization). 2013. *The Global Learning Crisis: Why Every Child Deserves a Quality Education*. Paris: UNESCO.

World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank.

World Bank. 2019. "Ending Learning Poverty: What Will It Take?" World Bank, Washington, DC. https:// openknowledge.worldbank.org/handle/10986/32553.

World Bank. 2020. *The COVID-19 Pandemic: Shocks to Education and Policy Responses*. Washington, DC: World Bank.

4. An evolving agenda for measuring the quality of service delivery

For the past decade, Service Delivery Indicators (SDI) surveys have provided evidence for policy makers, implementers, and communities seeking to improve the quality of health and education services in low- and middle-income countries. Now, COVID-19 (coronavirus) has dramatically raised the stakes in countries' efforts to strengthen the resilience of their health and education systems. Disruptions from the pandemic intersect with the longer-term demographic, epidemiologic, and technological trends that were already redefining health and education agendas and citizens' expectations for public services in many countries. Meeting the resulting challenges will require innovative policy and delivery solutions.

A new generation of SDI instruments is emerging to inform those solutions with a diversified repertoire of evidence.

The analysis in earlier parts of this book built on a core set of information common across all countries. This chapter looks toward the future of the SDI surveys, showing how the initiative's measurement work is evolving and diversifying to meet changing country needs. The chapter explores innovations in the measurement of frontline health and education service delivery. In doing so, it reflects changes in thinking about service delivery and measurement technology that have emerged over the past 10 years. The chapter first discusses adaptations that are enabling new SDI surveys to reflect specific country needs and policy priorities, while preserving the comparability of core indicators. It then describes new measurement approaches that better capture (1) the determinants of service provider performance, such as the quality of management, and (2) the influence of household contexts on health and education outcomes. The chapter's final sections discuss strategies for a more systematic uptake of health and education service quality measurement in policy and explore how the production and use of data on service delivery are changing in the wake of COVID-19.

Adapting SDI surveys to different country contexts

SDI core indicators are collected uniformly across countries, allowing benchmarking and highlighting areas for urgent action. At the same time, the SDI surveys are purposefully designed to allow flexibility in adapting to country-specific challenges. The interaction between national and cross-country experience generates a productive tension as well as innovations that have changed both *what* some SDI surveys measure and *how* they measure it. Recent developments include expanding the catalogue of clinical vignettes on which providers' knowledge is tested to capture country contexts; creating new modules that measure patient satisfaction; designing questionnaires that capture data on job satisfaction, mental well-being, and soft skills among health workers and teachers; and creating new modules that assess pandemic preparedness at the facility level.

Nutrition and stunting

As a measure of height-for-age, stunting is understood to be a marker of delayed growth, reflecting cumulative insults to a child's health and nutrition. Stunting continues to be a global problem, and, in countries such as Guatemala and Madagascar, which are undertaking SDI surveys, stunting

rates approach 50 percent. (USAID 2014, 2018). Stunting has many determinants, including household ("demand-side") risk factors, such as lack of safe water, sanitation, and hygiene; use of biomass fuels; and nonexclusive breastfeeding practices (Danaei et al. 2016). Just as important, but less discussed, is the role of health care providers (the "supply side") in detecting delayed growth, advising about healthful practices, and delivering high-impact interventions to prevent and manage malnutrition. Although most household surveys measure population-based risk factors and their associated burdens, no facility survey currently measures supply-side capacity to detect malnutrition comprehensively among patients and to deliver nutrition services.

The new generation of SDI surveys aims to fill this gap, working on two complementary fronts. First, the new surveys engage the facility side, assessing providers' ability to detect delayed growth and advise caregivers about appropriate nutrition practices. Health care providers are tested on their ability to detect stunting during a simulated routine checkup. As with other SDI clinical vignettes, an enumerator presents the basic details to the health care provider, who is then encouraged to ask follow-up questions. The health care provider should follow the World Health Organization (WHO) guidelines on integrated management of childhood illness (IMCI) to assess the child's symptoms, measure the child's growth, and decide on appropriate next steps. The health care provider should be able to detect various warning signs for malnutrition, and providers are graded on their ability to adhere to clinical protocols, including correct diagnosis and treatment.

To arrive at a holistic understanding of the determinants of child stunting, provider assessments should be matched with household data. Collecting household data is the second key dimension of some of the new-generation SDI surveys. To collect such data, households in Guatemala were sampled within the catchment area of a health facility, with oversampling of households with women of childbearing age who are more likely to have young children. Household questions focus on access (and barriers) to health services, behavior of caregivers and mothers regarding reproductive health, antenatal checkups, compliance with vaccinations, child-rearing practices, and nutrition for pregnant women and children. Given the important role of community health workers, the SDI survey in Madagascar is being adapted to include a separate module to measure their knowledge about antenatal care and maternal and child nutrition.

Because nutrition is important from both health and education perspectives, questions related to school feeding will be considered for addition to the education questionnaire and will be tailored to the context of each country. The questions to be considered may cover topics such as the following:

- Number and grade level of recipients enrolled in or receiving services through a program
- Delivery modality (breakfast, snacks, lunch, dinner, take-home rations)
- Frequency and duration of delivery (daily, only during the school year)
- Quality of food provided (fresh food, local sources, nutrition, fortified, supplements)
- Links with parents
- Program management (monitoring, implementation, compliance, financing, guidelines, targeting, other issues)

Mental health and noncommunicable diseases in primary care

Developed initially to measure service quality in Sub-Saharan Africa, most of the original SDI clinical vignettes reflect a burden of disease—mostly communicable—typical of low-income contexts in the past. As trends in disease burden have evolved globally and the geographic scope of the surveys expands, the SDI surveys have been modified to fit a broader range of potential health conditions, including noncommunicable diseases (NCDs). As such, SDI surveys have evolved to reflect the double burden of disease, which is becoming more common in low- and middle-income countries (Min et al. 2018). This process has begun in Armenia, Bhutan, Moldova, and Ukraine, where new SDI vignettes have been tailored to include cases of depression and hypertension.

The WHO's Mental Health Gap Action Programme (mhGAP) intervention guide provides recommendations for treating a variety of mental health disorders in nonspecialized health care settings (WHO 2019). It has been increasingly recognized that mental health disorders represent a large burden of disease, accounting for nearly 5 percent of disability-adjusted life years worldwide; that this burden is growing more quickly than other conditions; and that specialized care remains inaccessible to most patients. The mhGAP intervention guide aims to remedy this last point by promoting the delivery of pharmacological and psychosocial interventions for common mental health complaints in community health care settings. In line with this approach, the new SDI mental health vignette aims to measure capacity for mental health interventions at the primary care level. An enumerator presents the basic details to a health care provider, and the health care provider is then graded on her ability to ask relevant questions, rule out other potential diagnoses, and arrive at a probable diagnosis. She should offer counseling, pharmacological interventions if appropriate, and referral to a higher level of care.

Similarly, the increasing burden of NCDs in low- and middle-income countries has led to calls for integrating NCD prevention programs into primary health care (Kruk, Nigenda, and Knaul 2015). This integration can be done through risk factor management, such as smoking cessation and weight control

advice, and through improved care at the facility level, including improving the diagnosis of health problems in their early stages. To measure these capacities, the SDI survey has added a new clinical vignette on provider knowledge and management of hypertension, one of the most common NCDs. Hypertension, which can lead to serious health complications such as heart disease and stroke, is often underdiagnosed in low- and middle-income settings (Kim and Radoias 2018). The health care provider is graded on her ability to arrive at the correct diagnosis, to prescribe appropriate antihypertensives, and to offer behavior-change recommendations. This new vignette is intended to measure how well providers are able to screen, treat, and care for a common NCD that might be encountered in daily practice.

People-centered services

Although SDI surveys offer valuable insights in contexts where the supply of services is insufficient or of low quality, they lack demand-side information that would allow inferences, either correlational or causal, on how poor quality or deficits on the supply side affect the demand for and use of education and health services. Nevertheless, correlational evidence between observed or perceived poor quality of services and bypassing or underusing services has been noted in both the education and the health literature. For instance, research in Kenya finds that the quality of maternal and newborn care available to the poor is significantly worse than the care available to the better off (Sharma et al. 2017), whereas, in Tanzania, pregnant women bypass their nearest primary care facility to deliver at more a distant health facility, citing concern about the quality of care at government dispensaries and health centers (Kruk et al. 2009). Likewise, education studies in the United States (Chetty et al. 2011) and Israel (Angrist and Lavy 1999) have shown a strong negative correlation between class size in early primary grades and college attendance and student learning outcomes, respectively, suggesting that low-quality schooling has lingering effects on student performance and future enrollment.

Several ongoing SDI initiatives will help to address this methodological and evidence gap in both the education and the health fields. First, efforts are underway to merge data collected through the SDI education and health surveys in Kenya with geospatial data on cell phone connectivity, road infrastructure, nighttime lights, and population characteristics. Linking data will open new avenues for understanding the correlations between supply-side factors, population access, and other levers for comprehensive human capital reforms. Second, household surveys are also being considered as components of ongoing and future SDI surveys. Notably, the ongoing SDI health survey in Guatemala incorporates such a household component.[1] Individuals living within the catchment area of a health facility who are included in Guatemala's SDI survey will

receive a household survey with questions focusing on access (and barriers) to health services. Third, the next generation of survey tools will feature a "people-centered approach" to measurement, soliciting input from service users (patients and schoolchildren or parents) on perceived quality, current patterns of service use, and intended future use. These data, joined to those obtained on the supply side, will enable future exploration of links between the provision of services and their use.

The new-generation SDI surveys are doing more to capture people's experiences with health and education services. For the health survey, this effort means measuring patients' experience of care. Patient exit interviews are aligned to the concept of a patient-centered approach. They ask patients whether a provider has engaged in respectful practices, including greeting them, introducing himself or herself, and asking if the patient has any questions. In addition, the patient exit interview collects information on whether the provider has explained the course of care, clarified the potential side effects of medications, and promoted the mitigation of risk factors or lifestyle changes.

In the revamped patient exit interview, information on both experience and satisfaction with care is solicited from respondents. Regarding their experience, patients are asked questions on wait time, visit duration, out-of-pocket costs of care, comprehension of care and follow-up procedures, and related topics. Patients are also asked to assess their satisfaction with the care they received during the visit. The survey aims to measure patient satisfaction across multiple domains, including respect and privacy, autonomy, and communication. To account for subjectivity in patients' assessment of satisfaction with care, new features of the questionnaire include anchoring vignettes that allow quantification of respondents' expectations and thus adjustment for subjectivity (Larson et al. 2019 and King et al. 2004). Collectively, the measurement of patient experience and satisfaction with care as provided in the exit interview can deliver more comprehensive and actionable evidence on improving patient centeredness in health systems.

The analysis in chapter 3 suggests that school characteristics cannot explain a large share of the variation in student learning. In order to understand factors associated with this unexplained variation, novel household and child questions are being explored with the objective of capturing the experience of students and families with education systems and the determinants of education investments. Among other changes, survey designers are considering new sections on how parents and children perceive the quality and skills of teachers, including emotional support skills. For instance, parents are being asked to say what they think makes teachers good at improving student learning and to compare two teachers in different scenarios and choose which one they would like for their children.

There is growing evidence that parental beliefs about their children's learning process and expectations about the returns to education can shape parental

investment (Bergman 2021).[2] Questions on parents' expectations as well as their perceptions of the strengths and weaknesses of their own children with regard to learning have also been included in recent SDI surveys. For instance, parents are being asked what level of education they expect their child will be able to complete and how they think their child is performing compared to classmates. An additional section asks parents a battery of questions about whether they think that their child is considerate of other people's feelings, shares readily with others, has temper tantrums, worries too much, and similar issues. Understanding how beliefs and expectations are correlated with student learning and school choice is an important new area of research.

The World Bank recently launched the Global Education Policy Dashboard, which integrates key SDI indicators into a systemwide framework that also includes essential elements from the Systems Approach for Better Results (SABER), the Measuring Early Learning and Quality Outcomes (MELQO) project, the World Management Survey, the Bureaucracy Lab surveys, and others. The SDI team and Global Education Policy Dashboard team are leveraging synergies on data collection, technical expertise, and innovations for countries' benefit. For instance, current SDI surveys do not capture proxies for school readiness, an important determinant of learning in school. The Global Education Policy Dashboard has developed a cost-effective way to assess student vocabulary in first grade that the SDI team will build on in the future.

Provider job satisfaction

The health care provider questionnaire has also been expanded to capture the experience of employees, including questions on work satisfaction, workload, stress management, and challenges. The questions on employee satisfaction are based on multiple surveys, including the Satisfaction of Employees in Health Care (SEHC), which has been validated in low- and middle-income settings (Chang et al. 2017). The workload section is designed to measure the burden of both clinical care and administrative tasks, responding to concerns about the overburdening of health care workers. Finally, the survey also includes questions related to facility management, such as satisfaction with pay scales, promotion systems, supervision, and opportunities for career advancement. Overall, the health care provider questionnaire is intended to describe the work environment and to highlight areas where it could be improved.

In education, SDI surveys have been adapted to collect data on mental well-being (particularly in fragile and conflict contexts) and on socioemotional skills among teachers and parents, in line with the growing evidence on the importance of these elements for children's learning. For instance, in addition to the standard SDI modules, the SDI education survey in Indonesia (2019) includes a teacher module on socioemotional skills and a household module

aimed at assessing parental school choice and socioemotional conditions. Both the teacher and parent modules capture self-reported measures of socioemotional skills and include sections with measures derived from the Item Grit Scale, the Big Five Inventory, and the Growth Mindset.[3] Preliminary results will be shared in the forthcoming Indonesia SDI report 2021.[4]

Emergency preparedness

COVID-19 has highlighted the importance of preparedness for pandemics and other large-scale emergencies. Adequate preparedness and effective response may help to reduce the transmission of a pathogen; limit the number of cases, hospitalizations, and deaths; maintain essential services not related to the outbreak; and reduce an epidemic's economic and social impact. Rational personnel management, optimal use of medicines and other supplies, and careful triaging of patients are all important to ensure that health services are effective in responding to emergent health challenges.

However, evidence to aid comprehensive pandemic planning is scarce. In many countries, concerns have emerged about whether frontline health facilities have adequate patient care and infection control capacities to protect patients and health workers during an outbreak. In the absence of reliable, real-time data on these indicators, large-scale health facility surveys like the SDI surveys can be used to gauge health system preparedness, as has been done for the COVID-19 pandemic (Sharma et al. 2020). In addition, SDI surveys have begun to include more measures of preparedness, with a new preparedness module incorporated in the forthcoming SDI survey in Bhutan, for example. Measuring preparedness can involve surveying facilities to determine whether they have an emergency plan; whether appropriate contingency mechanisms exist for the isolation, transport, and treatment of contagious patients; and whether they have surge capacity.

The COVID-19 pandemic and the associated lockdowns have brought a renewed focus on providing continuity in learning when children cannot be physically in school. Students' and teachers' digital skills, school connectivity, and information and communication technology (ICT) infrastructure need to be seen (and measured) in a new light. Even before the pandemic, the World Bank Group's SABER–ICT Initiative and the United Nations Educational, Scientific, and Cultural Organization's Institute for Statistics were seeking to improve the availability of policy-relevant data, information, and knowledge on what matters most in using ICT to improve the quality of education. COVID-19 has strengthened the importance of ICT and multiplied the channels through which it can affect learning. Teachers and students now need to be empowered to use technologies to continue engaging in learning activities as schools slowly reopen.

All schools, including the most remote, should have access to ICT, both to teach students how to use these tools and to connect with a wider array of

learning materials. Likewise, school principals need to coordinate activities and connect with teachers and students virtually. Existing survey tools will need to be adapted further to capture these and other novel aspects of ICT that can shape learning outcomes. With this in mind, the World Bank Group's Education Global Practice has been developing the EdTech Readiness Index, composed of eight indicators, including connectivity, teacher training, digital learning resources, and online assessments. This initiative and the SDI team's ongoing collaboration with the Education Global Practice could create additional synergies in the future. Similarly, to the extent that COVID-19 itself continues to pose a challenge, new physical requirements for spaced seating; water, sanitation, and hygiene infrastructure; and protective equipment will become essential elements of school life.

Addressing determinants of provider performance

The performance of individual service providers (such as frontline health care workers and teachers) is a crucial determinant of the quality of services that citizens receive. What factors shape how well frontline service professionals perform, and what levers exist to support providers in doing their best work? SDI instruments are evolving to deliver new evidence on why providers make certain good or suboptimal choices in their daily practice and what the role of management is.

Better understanding providers' performance

In health, the current SDI vignettes capture whether a provider gives the correct diagnosis and treatment, and are being expanded to provide more information on why a provider is diagnosing a condition incorrectly. For example, although the existing data show that many providers are not effective at diagnosing diabetes, it is not clear whether they think that the patient is healthy or instead is displaying risk factors for another disease, such as depression or chronic fatigue, that might present similarly to diabetes. At present, health care providers can generally diagnose the tuberculosis vignette quite well, but they do poorly with treatment. Based on consultation with medical experts, the questionnaire now includes more of the most common incorrect answers as well as the requirement to record any nonstandard answers given. This information will shed light on the most common errors and will help to provide more pinpointed recommendations to improve provider performance.

In schools, teacher observation captures more granular information about time spent on different activities by a teacher and students during a class

period. The SDI methodology that is used to gather this information closely follows the widely applied Stalling's Classroom Observation System, in which a 30- to 60-minute period is observed and, for each minute, the enumerator notes a "snapshot" of current activities in the classroom. The instrument allows an enumerator to record what every person in the classroom is doing, including whether they are engaged in a learning activity or not. The information gathered is used to construct the time-on-task indicator, which refers to the amount of time during a class period in which a teacher is actively engaged in teaching or students are actively engaged in assigned learning tasks. Analysts can combine this indicator with data from the staff roster module (used to measure absence rate), the classroom observation module, and reported teaching hours to estimate the time spent teaching per day.

Other classroom observation tools that take advantage of new technologies have been used in more recent SDI surveys. For instance, the TEACH tool, developed by the Education Global Practice, was tested for the Mozambique SDI survey in 2018, in place of the standard classroom observation module. TEACH is similar to the SDI tool but generally involves recording classes on video and then having a trained observer view the videos and provide an assessment.[5] In addition to time on task, this approach assesses the quality of teacher practices through behavioral information observed and classified into three categories: classroom culture, instruction, and socioemotional skills. Recent efforts on this front are trying to leverage machine-learning techniques to identify teacher-student interactions and engagement automatically through data mining of videos of classroom observations (see, for instance, Aung, Ramakrishnan, and Whitehill 2018). If proven reliable, these methods might help to reduce the cost of data collection and increase the information available to inform policies that promote effective teaching and learning in classrooms.

The role of management

Limited competence and high absenteeism among service providers can reflect many underlying factors, including the incentives that agents face within an organization, the design of which can have both positive and negative effects on the motivation and effort of employees (Ashraf, Bandiera, and Jack 2014; Björkman Nyqvist et al. 2019; Karachiwalla and Park 2017; Rasul and Rogger 2018). Intrinsic motivation also matters for a provider's performance (Ashraf, Bandiera, and Lee 2018; Deserranno 2019), as do the financial resources that organizations and staff must work with and the selection of staff members in the first place (Das et al. 2013; Deserranno 2019). Together with these factors, which interact with each other (Donato et al 2017), growing empirical evidence shows that management practices have an important role to play in how organizations perform (Bloom, Sadun, and Van Reenen 2016).

The explanatory power of management carries over into the space of public service delivery. For example, Rasul and Rogger (2018) find that bureaucratic autonomy in Nigeria increases project delivery rates (although increasing incentives and monitoring among bureaucrats has the opposite effect). The quality of management (such as organizational practices in schools) has also been shown to predict educational outcomes (including math or other assessment results) within countries such as Brazil, Canada, India, Italy, Sweden, Uganda, the United Kingdom, and the United States (see Bloom et al. 2015; Crawfurd 2017; Di Liberto, Schivardi, and Sulis 2015). Interventions aimed at improving the quality of management in schools have also been shown to improve student performance, although the effectiveness of such measures varies by context (see Blimpo, Evans, and Lahire 2015; Fryer 2017).

The recent education SDI survey in Indonesia was coupled with the implementation of the Development World Management Survey (D-WMS) among school directors to assess the quality of school management. The D-WMS tool was created in 2008 as an adaptation of the original World Management Survey and covers 15 questions across the following five main areas: leadership, operations management, performance monitoring, target setting, and people management. Through an interview-style questionnaire aimed at school administrators (either principals or their second-in-command), trained and certified enumerators assign scores ranging from 1 to 5 along the five main areas. Again, preliminary results are included in the 2020 SDI report for Indonesia, but, because of the richness of the data, their link to the quality of service provision and student learning will be analyzed further in a separate publication. This methodology will continue to be replicated in future SDI surveys whenever possible.

Existing evidence suggests that management practices also affect quality of care in the health sector (Macarayan et al. 2019). Effective health care management requires oversight of many facets of service delivery, including the facility's layout and patient flow, development and standardization of protocols, staff recruitment and retention, effective use of personnel and equipment, setting of targets, continuous tracking of performance, and use of accountability systems with performance incentives. Although many of the inputs into the provision of quality care, such as provider training or medical supplies, may be costly, management practices can often be improved in situations with constrained resources. When properly implemented, management interventions have the potential to be cost-effective.

A forthcoming survey in Kenya will contribute to the existing, though somewhat limited, evidence on the role of management practices in improving the quality of service delivery in health facilities. Specifically, a telephone survey will be implemented to assess the quality of management in a sample of hospitals and lower-level primary health care facilities that were included in Kenya's 2018 SDI survey. The sample includes both public and private

facilities, enabling an examination of differences between management practices and incentive systems by facility ownership. The survey builds off two highly used management tools—the D-WMS and the SafeCare Standards—and was adapted to fit the Kenyan primary health care context.[6] Five hundred facility managers will each participate in an hour-long phone survey, in which they will be asked a series of validated questions covering 17 domains of management practices, including leadership vision and goals, standardization of protocols, performance tracking and review, and continuous improvement. This survey not only will allow for benchmarking of management practice outcomes in Kenya against practices in other settings (given that the D-WMS methods are standardized) but also will introduce innovations to measure management practices relating to staff absenteeism and stockouts of equipment, supplies, and medicines.

Beyond management, various SDI surveys have also attempted to get a general idea of the revenue composition of both schools and health facilities—that is, whether they finance their operations with their own resources, cost recovery, transfers from the central level, transfers from decentralized levels, donations, or any other means. Although these data have proven to be extremely useful for understanding the potential implications of education and health financing reforms at the country level, their collection and cross-country comparability have been challenging because of country-specific financing and organizational structures, as well as data privacy concerns. Likewise, having a broad understanding of the composition of expenditures (human resources, equipment, and maintenance, among others) would be invaluable for understanding differences in efficiency between facilities and over time. As the SDI program continues to evolve, it will inevitably grapple with the balance between cross-country comparability and country-specific needs, particularly as they relate to financing, structural organization, and expenditures.

Understanding interactions with family background

Learning, child development, and health are outcomes of complex processes that are based on access to quality schooling and affordable, quality health care, but also on individual and family choices. A large body of evidence shows that children in poor households experience worse human capital outcomes, and several factors can contribute to these gaps. Lack of financial resources and limited ability to borrow often prevent poor households from accessing health and education services. External shocks may force children in poor households to drop out of school or may otherwise adversely affect their learning outcomes.

Households may face significant barriers or opportunity costs in acquiring human capital. In addition, social norms about women's roles, which may be related to socioeconomic status, shape many critical decisions related to human capital, such as fertility, breastfeeding, or schooling (World Bank 2019).

Understanding the individual and family characteristics that are likely to shape children's learning and health is an important complement to understanding service use. In education, older surveys collected only basic information on family characteristics, such as whether a child had breakfast that morning. More recently, SDI surveys in both health and education have expanded these efforts. For instance, the SDI education surveys in the Democratic Republic of Congo (forthcoming 2021) and in Indonesia (2019) collected data using household modules from a subsample of students tested in SDI schools (Yarrow et al. 2020). These modules include questions on topics such as parental education, household characteristics (infrastructure, work and other sources of income, and household structure), assets (both agricultural, such as livestock and crops, and nonagricultural, such as ownership of a television, radio, computer, motor vehicle, or refrigerator), books at home (number and type), and type of educational expenditures (school fees, food, and others). Likewise, the forthcoming SDI health survey in Guatemala includes a household module that collects information about household characteristics, health expenditures, and demand for health services. These new initiatives will shed light on potential improvements in measurement for future SDI surveys and yield additional evidence on the link between household characteristics, demand for services, and quality of service delivery.

More innovative methods to connect with households also show promise and can inform the evolution of SDI methodology in the future. For example, in the SABER–Service Delivery survey in Punjab, Pakistan (2018), parents' phone numbers were obtained from school records, and a phone interview was conducted to document household characteristics. This information can be linked to children's learning outcomes to explore differences by socioeconomic status. Recent evidence from the LEAPS study, also in Punjab, Pakistan, shows that information on household assets and parental education collected from students at the school level correlates well with information collected from parents at the household level during an in-person visit (Bau, Das, and Yi Chang 2021).

Finally, the analysis of health care in this book offered a preview of how the combination of SDI data and administrative data can help to identify bottlenecks in the delivery of health and education services. There may be many reasons behind drug stockouts in health centers and the absence of soap in school bathrooms, including lack of road access or inadequate power supply. Combining SDI data with administrative data and geocoded information indicators (such as road access, electrification, and water supplies) can help to improve the accountability of government overall, not only in the health and education sectors.

Improving comparability of SDI surveys over time

During the past decade, SDI surveys have evolved steadily. Each survey has been tailored to address specific policy and research questions in its setting of implementation. Although such specificity means that the surveys are well designed to respond to pressing national policy concerns, it also poses challenges for the direct comparability of some surveys from the same country over time. In addition, the SDI surveys are designed to provide just-in-time findings in fragile and conflict-affected settings, which means that logistical and safety considerations may also affect comparability. Experiences from multiple surveys in the same country are now being used to inform the design of future repeated surveys. This section outlines challenges, solutions, and lessons learned from repeated SDI surveys in Kenya, Mozambique, and Tanzania.[7]

The methods used in the Tanzania SDI education surveys in 2014 and 2016 allow for direct comparisons across the two years. Over this period, student learning outcomes for fourth-grade students, as measured by the English, math, and Swahili SDI tests, improved. The mean student test score across all subjects improved by an average of 11 percentage points. Other SDI indicators for which a significant change was observed in this two-year period included the share of students with pens, pencils, and an exercise book (increased from 84 percent to 92 percent) and the proportion of schools with a functioning blackboard (increased from 74 percent to 83 percent). Interestingly, the proportion of students with a textbook declined significantly (from 25 percent to 19 percent), while the observed student-teacher ratio increased (from 43 percent to 47 percent), perhaps indicating that system resources have not kept pace with enrollment. Smaller changes were observed in other SDI indicators, such as teachers' absence from the classroom (declined by 5 percentage points) and teachers' subject knowledge in language (deteriorated slightly) and mathematics (increased significantly, but only on certain tasks, such as Venn diagram and graph interpretation). Pedagogical knowledge also improved modestly. Trends in basic infrastructure showed no noticeable improvement.

The 2014 and 2016 education surveys for Tanzania are so far the only SDI surveys that track a complete panel of schools over time. The direct comparability of data is the result of a careful panel sampling strategy and the use of identical survey tools. These methodological and logistical choices allow for accurate measurement of changes over time to track progress in education outcomes and process indicators. Nonetheless, the trends observed in Tanzania also suggest that, whereas certain indicators can be improved substantially in a two-year period, others may take longer. In addition, the measurement error associated with some indicators may make direct comparisons harder to interpret. For instance, many factors may affect the performance of students and

teachers in SDI assessments, including how they are feeling on the day of the exam, whether they had breakfast, their environment, and even the weather. In contrast, directly observing whether there is a blackboard or bathroom in a school is subject to much less measurement error. In the future, teams interested in carrying out repeated SDI surveys will need to identify a relevant interval between surveys, consider specific project needs, and understand which indicators are likely to show meaningful changes over the defined period.

Mozambique implemented repeated SDI education surveys in 2014 and 2018. Over this longer interval, researchers observed significant changes in both student knowledge (increased from 21 percent to 31 percent of correct answers) and teacher knowledge (increased from 31 percent to 42 percent of correct answers). At the same time, whereas teacher absence was greatly reduced (from 45 percent to 30 percent), the percentage of schools with minimum equipment dropped significantly (from 77 percent to 68 percent). Differences in all of the remaining SDI indicators are less than 2 percentage points across both years. The group of variables that exhibit significant differences is similar to that in Tanzania. Some level of caution is warranted in directly comparing the two sets of SDI results from 2014 and 2018 in Mozambique, because logistical factors resulted in important differences in sampling. Among other limitations, ongoing conflict made some regions impossible to survey in 2018, so these areas were excluded from the sampling process. Additionally, in 2018, the survey team faced several logistical constraints that prevented enumerators from reaching all schools in the sample and from carrying out the second (unannounced) visit in many others. Moreover, fieldwork for the 2014 survey started approximately three months earlier in the calendar year than the fieldwork in 2018. For this reason, data were collected at different moments of the school year, which could affect (1) student learning, because students in one survey had more schooling, and (2) absence indicators, because both teachers and students may be more likely to be absent as the school year draws to an end, for example. Although the final analysis employed adjusted weights and explored differences using various sensitivity analyses and scenarios, the magnitude of the differences observed could be attributed partially to some of the above circumstances. For instance, one sensitivity analysis suggested that teacher absence for 2018 was between 30 percent and 40 percent, which would amount to a reduction of anywhere between 5 percentage points and 15 percentage points, relative to 2014. Lessons learned from the repeated Mozambique surveys underscore the challenges of maintaining comparability over time, particularly during periods of conflict, and emphasize the need for additional research on the role of seasonality in education survey findings.

In Kenya, an SDI health survey was conducted in both 2012 and 2018. As discussed in the 2018 report for Kenya (World Bank and Government of Kenya 2019), important differences between the 2012 and 2018 survey rounds may

influence observed differences in results. Comparing the values for 2018 with those for 2012 shows that Kenya appears to have experienced an increase in caseload, absence rate, and availability of some infrastructure. Although the two survey rounds used very similar instruments, methodological improvements and differences in the nature of the samples led to differences in how sampling weights are applied in calculations using data from the two rounds. Because of these factors, differences between the 2012 and 2018 values, particularly for indicators that use provider-level disaggregated data, may reflect differences in weighting schemes rather than true underlying changes.[8] The challenges in generating comparable results between the two surveys underscore the trade-off between constantly improving the survey and its methods and ensuring that policy-relevant changes over time can still be examined. In the future, SDI methods will increasingly emphasize gathering and retaining details on survey sampling frames, which can help analysts to adjust for sampling differences and improve comparability.

Results in action: How SDI surveys inform program operations

In a decade of implementation, the SDI surveys have had a significant impact in countries. The surveys serve multiple purposes for policy makers and program implementers. They can help to provide insight into aspects of service delivery that are typically not well measured. For known health or education challenges, they can help to diagnose root causes. From an operations standpoint, they can help governments and the World Bank to measure the success of projects over time. Finally, they can help to raise awareness of the quality of service provision and thereby stoke demand to implement changes. Ideally, an SDI survey will do all of the above, but a few examples are useful to show how each of these changes has happened in the real world.

For example, in the Democratic Republic of Congo, a 2019 education survey is providing an understanding of several aspects of service delivery quality, including infrastructure, learning materials, teacher practices, student learning, and sector governance. Most of these indicators are not yet captured by the country's nascent education management information system, and the SDI data allow government oversight in areas with otherwise little to no information. As a result, the SDI survey has informed and strengthened a World Bank operation. Several measures from the SDI survey are being used as indicators for the Education Quality Improvement Project (EQUIP), whose objectives are to strengthen the quality of learning conditions in primary schools

and improve education sector management. The SDI survey was initially conducted in a representative sample of more than 1,500 public primary schools and will be repeated again in the same sample of schools in 2022. Having a baseline and endline will allow the World Bank team to assess the project's progress toward its objectives.

In Madagascar, SDI has been instrumental in offering fresh solutions to long-standing problems. Madagascar is among the countries with the highest rates of chronic malnutrition, a problem with potentially disastrous long-term impacts on the economy. A health SDI survey in 2016 revealed a low share of clinicians with the capacity to diagnose and treat children and to manage maternal and neonatal complications. In this case, although the problem of malnutrition was well understood, the SDI survey indicated the need for better training of health care providers to manage these. In this respect, SDI evidence directly informed the design of the Improving Nutrition Outcomes Project, which aimed at increasing the diagnosis and treatment capacities of frontline health care providers.

SDI surveys have also helped to raise public awareness of service quality. As an example, the 2011 SDI survey in Tanzania showed high rates of teacher absence, among other indicators. This finding led to a strong public response and an increased demand for teacher competence and presence in public schools. Building off this demand, the government introduced the "Big Results Now!" Initiative, with the goal of strengthening accountability, improving incentives, and providing better support for students. The US$416 million program was followed by assessments to measure students' reading and math skills, which showed improved learning outcomes over the ensuing years. A similar story took place in Madagascar, where the 2016 SDI survey highlighted high absence rates among school directors and a deficit of necessary content knowledge and pedagogical skills in teachers. Those results helped to inform preparation of the country's Education Sector Plan for 2018–2022, a bold and credible plan to improve the quality of learning, access, and governance. Results like this show how SDI surveys can highlight opportunities for reform and catalyze political momentum to make change happen.

Reinventing measurement after COVID-19: Using familiar tools in new ways

In addition to promoting evidence-driven delivery reforms in countries, SDI research contributes to a global agenda of measurement practice that is evolving quickly in the wake of COVID-19. COVID-19 is dramatically reshaping service delivery in many countries and posing new challenges to measurement. Focus in many settings has turned to remote data collection. Although this approach

has been adopted by necessity in the short term, it also offers promise for future practice. Data collection via phone surveys cannot always match the depth of in-person data collection—for example, unannounced visits to measure teacher absence or visual verification of the stock of medicine in a clinic or textbooks in a classroom. However, phone surveys offer a useful and timely way to generate more data on key services.

As countries respond to the COVID-19 pandemic, health systems face the simultaneous challenge of maintaining continuity in essential health services. Identifying which services are disrupted, quantifying the level of damage, and matching these estimates with knowledge of the local context can inform a more effective public health response. In Guatemala, up to four rounds of periodic phone surveys will be collected before a future in-person SDI survey, generating a panel data set that will guide the government's response to the pandemic and provide a unique opportunity to study time trends in service provision in the country during a prolonged emergency. The phone survey instrument being used in Guatemala[9] covers the following aspects of essential health services (WHO 2020):

- Health workforce (numbers, absences, COVID-19 infections, training, and support)
- Financial management and barriers
- Service delivery and use (facility closures, community communication campaigns, changes in service use, and catch-up strategies)
- Infection prevention and control capacities (protocols, safety measures, guidelines, and availability of personal protective equipment for staff)
- Availability of therapeutics, diagnostics, and supplies and vaccine readiness
- Provision of COVID-19–related services

Phone surveys are relatively inexpensive, allowing facilities to provide direct feedback on challenges in providing essential health services. They are also timely, collecting additional information in the context of identified disruptions, on which governments can act quickly. However, such surveys are not a silver bullet. For example, higher rates of nonresponse, particularly in areas with poor phone connectivity and in facilities already overstretched because of COVID-19, might prevent the health facilities in greatest need from participating in telephone surveys. Additionally, the nature of a phone survey limits the amount and type of information that can be collected, although such shortfalls could be partially addressed by triangulating survey data with administrative data to improve overall data quality.

The experience of delivering schooling and care remotely during the pandemic has permanently changed many aspects of service delivery. As countries build back better, measurement will need to reflect these changes, such as the interaction of teachers and students with technology, the experience of patients with telemedicine, and the effect of these changes on the quality of service.

Using measurement to inform policy dialogue and strengthen country capacities

By measuring how services are delivered, results from SDI surveys can help to link resources to results and shift national policy dialogue from inputs toward results and quality. SDI surveys provide a platform to trigger policy debates and reforms. This and preceding chapters have discussed specific examples. In Mozambique, a national campaign to reduce teacher absence was instituted after the launch of the 2014 SDI education survey report (Bassi, Medina, and Nhampossa 2019). Togo's SDI education findings sparked a revision of teachers' college curricula and a novel collaboration between a teachers' union and the education ministry to deliver regional workshops aimed at better learning outcomes. In the Democratic Republic of Congo, SDI surveys were incorporated into an education quality improvement project. In Tanzania, as this book has detailed, SDI data were used as diagnostics during the planning of major reforms, and indicators from SDI were added to the Big Results Now! in education monitoring framework. Policies based on SDI evidence resulted in more equitable deployment of teachers within and across districts in Tanzania, and teachers were motivated to spend more time in classrooms teaching.

Inherent in the design and implementation of the SDI surveys is extensive capacity building. Given that the surveys must be adapted to fit each country's specific context, staff from the ministries of health and education, other local stakeholders, development partners, and research groups are intimately involved in each step of the work. The SDI model is to contract and empower local groups to implement the study, meaning that the survey is enriched by contextual expertise, while local experts are deepening their skills in survey methodologies. A standard part of each survey activity is capacity building for local analysts. This capacity building has been accompanied by larger training workshops, held so far in Kenya, Nigeria, and Tanzania, where policy analysts, researchers, medical statisticians, and civil society activists from countries around the region have received both basic and advanced training in the use and application of data.

Future measurement initiatives have the potential not only to reinforce national capacity but also to improve administrative data systems. In many settings, surveys would not be needed to assess many aspects of service delivery, if national administrative systems were able to provide timely and accurate information. However, in many low- and middle-income countries, this is not yet the case. *World Development Report 2021: Data for Better Lives* analyzes the many stubborn barriers to achieving well-functioning national administrative systems (World Bank 2021). The SDI surveys are intended to build on routine data systems, allowing for triangulation and verification. At the same time, the SDI surveys are designed to be a separate and objective measurement, providing an independent measure of services. Overall, measurement work of any kind may

contribute to strengthening national administrative systems by demonstrating and promoting the use of data for policy making; enhancing local capacity to produce, use, and reuse data; and helping to build a proper legal and regulatory framework for data generation and applications.

The evolving scope of the SDI surveys, as outlined in this chapter, will allow for deeper analysis and a greater range of analytic objectives. However, these new initiatives will need to be balanced against the increased complexity and cost of the survey. Part of the goal of these revisions has been to increase the flexibility of the SDI survey, allowing for deep dives into issues that interest the national government, while maintaining a core of comparable measurements across countries. The SDI survey can be woven into existing country-level data initiatives, adding detail where needed. Because countries vary in their burden of disease, the organization of their health systems, and their existing data systems, any additions will be context dependent and tailored to the specific objectives of the survey.

A wider perspective: Measurement as a public good for research

The SDI evidence is intended to guide policy makers, stakeholders, and citizens in shared action to strengthen health and education systems—enabling services to work better for average people and laying the foundations for inclusive growth. Although this contribution is important, the potential uses of SDI data do not end there. In addition to the direct usefulness of its results for policy makers, service providers, and citizens, the SDI initiative provides relevant inputs for future research.

The SDI program has invested considerable resources in cleaning, harmonizing, and anonymizing SDI data to make the information available and useful to the scientific community. SDI data are harmonized and available through the data portal, www.sdindicators.org. The earlier releases of some of these data were accompanied by a significant increase in the number of publications using SDI data or referencing the SDI survey tool. Thus far, SDI surveys have been featured in about 45 publications authored by local and international scholars. Among these publications, more than 15 peer-reviewed academic articles have been published in leading economics, education, and health journals.[10] Use of the data helps to underline the surveys' rigor and relevance for both policy and research.

The SDI program expects that the most recent data release, with improved harmonization methods, will make the data more accessible to researchers and promote their use in a growing number of academic publications. In particular, the time trends in countries with repeated SDI surveys (such as Kenya and Tanzania) offer important opportunities for further learning.[11] Simultaneously, this new release seeks to facilitate the work of policy-oriented

organizations by making the data more usable and comparable across countries and over time.

The future growth of SDI surveys will depend on increased engagement and buy-in from country governments, World Bank teams, and other partners. The SDI surveys offer potential lessons for every country and can inform many local policy processes. At the same time, to serve optimally as a global good, the surveys should also expand in both their frequency and their coverage. Increasing outreach and engagement (including dissemination, research, and publication of case studies) can help to ensure growing demand for SDI surveys in the years ahead.

Notes

1 | This survey is on hold because of the COVID-19 pandemic but will resume as soon as conditions permit.

2 | Attanasio and Kaufmann (2014) provide detailed discussions of these and related topics.

3 | As measured with four items adapted from Blackwell, Trzesniewski, and Dweck (2007).

4 | These data require further independent analysis to understand (1) whether these measures are capturing the desired skills and domains in low- and middle-income countries, despite being created and calibrated in high-income countries, and (2) whether these measures correlate with student learning results. Future SDI surveys will continue to draw on this fast-evolving and important literature. Use of these approaches has been shown to correlate with improvements in learning outcomes and overall student well-being in some instances.

5 | The classroom observation can also be done in person, if needed. This methodology was used in Jordan, for instance. Sometimes, more than one observer can provide assessments, and scores are then reconciled.

6 | World Management Survey, accessible at https://developingmanagement.org. SafeCare Standards, accessible at https://www.safe-care.org/who-we-are/safecare-standards/.

7 | The lessons and statistics in this section are drawn from the analytical results of Trako, Molina, and Asim (2019) for Tanzania; Bassi, Medina, and Nhampossa (2019) for Mozambique; and World Bank and Government of Kenya (2019) for Kenya.

8 | Specifically, the 2018 sample was much more comprehensive: more than 10 times as many facilities and almost 9 times as many providers were surveyed in 2018 as in 2012. The 2012 survey was conducted in 15 counties and is representative only at the national level, whereas the 2018 round was conducted in 47 counties and is representative at both the national and the county levels. Unlike the 2018 survey, the 2012 survey did not include for-profit private facilities.

9 | This instrument was designed by the World Bank Group's Human Development and Public Services team at the Development Research Group and the Global Financing Facility for Women, Children, and Adolescents, with the support of the Primary Health Care Performance Initiative and SDI teams.

10 | To date, these journals include the *Review of Development Economics and Social Science and Medicine*, the *International Journal of Hygiene and Environmental Health*, the *Journal of Economic Perspectives*, and others.

11 | This book focuses primarily on within- and across-country comparisons. Only the latest survey results have been included for countries that have implemented multiple rounds of SDI surveys. As discussed, methodological improvements in the surveys over time make temporal comparisons complex and require a more extensive analysis that is beyond the scope of this book.

References

Angrist, J. D., and V. Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114 (2): 533–75. https://doi.org/10.1162/003355399556061.

Ashraf, N., O. Bandiera, and B. K. Jack. 2014. "No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery." *Journal of Public Economics* 120 (December): 1–17.

Ashraf, N., O. Bandiera, and S. Lee. 2018. "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services." London School of Economics and Political Science, LSE Library, London.

Attanasio, O. P., and K. M. Kaufmann. 2014. "Education Choices and Returns to Schooling: Mothers' and Youths' Subjective Expectations and Their Role by Gender." *Journal of Development Economics* 109 (July): 203–16.

Aung, A. M., A. Ramakrishnan, and J. R. Whitehill. 2018. "Who Are They Looking at? Automatic Eye Gaze Following for Classroom Observation Video Analysis." International Educational Data Mining Society.

Bassi, M., O. Medina, and L. Nhampossa. 2019. "Education Service Delivery in Mozambique: A Second Round of the Service Delivery Indicators Survey." World Bank, Washington, DC.

Bau, Natalie, J. Das, and A. Yi Chang. 2021. "New Evidence on Learning Trajectories in a Low-Income Setting." *International Journal of Educational Development* 84 (July): 102430.

Bergman, P. 2021. "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment." *Journal of Political Economy* 129 (1): 286–322.

Björkman Nyqvist, M., A. Guariso, J. Svensson, and D. Yanagizawa-Drott. 2019. "Reducing Child Mortality in the Last Mile: Experimental Evidence on Community Health Promoters in Uganda." *American Economic Journal: Applied Economics* 11 (3): 155–92.

Blackwell, L. S., K. H. Trzesniewski, and C. S. Dweck. 2007. "Implicit Theories of Intelligence Predict Achievement across an Adolescent Transition: A Longitudinal Study and an Intervention." *Child Development* 78 (1): 246–63. http://dx.doi.org/10.1111/j.1467-8624.2007.00995.x.

Blimpo, M. P., D. K. Evans, and N. Lahire. 2015. "Parental Human Capital and Effective School Management." Policy Research Working Paper 7238, World Bank, Washington, DC.

Bloom, N., R. Lemos, R. Sadun, and J. Van Reenen. 2015. "Does Management Matter in Schools?" *Economic Journal* 125 (584): 647–74.

Bloom, N., R. Sadun, and J. Van Reenen. 2016. "Management as a Technology?" NBER Working Paper W22327, National Bureau of Economic Research, Cambridge, MA.

Chang, E., J. Cohen, B. Koethe, K. Smith, and A. Bir. 2017. "Measuring Job Satisfaction among Healthcare Staff in the United States: A Confirmatory Factor Analysis of the Satisfaction of Employees in Health Care (SEHC) Survey." *International Journal for Quality in Health Care* 29 (2): 262–68.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126 (4): 1593–660.

Crawfurd, L. 2017. "School Management and Public-Private Partnerships in Uganda." *Journal of African Economies* 26 (5): 539–60. https://doi.org/10.1093/jae/ejx021.

Danaei, G., K. G. Andrews, C. R. Sudfeld, G. Fink, D. C. McCoy, E. Peet, A. Sania, et al. 2016. "Risk Factors for Childhood Stunting in 137 Developing Countries: A Comparative Risk Assessment Analysis at Global, Regional, and Country Levels." *PLoS Medicine* 13 (11): e1002164.

Das, J., S. Dercon, J. Habyarimana, P. Krishnan, K. Muralidharan, and V. Sundararaman. 2013. "School Inputs, Household Substitution, and Test Scores." *American Economic Journal: Applied Economics* 5 (2): 29–57.

Deserranno, E. 2019. "Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda." *American Economic Journal: Applied Economics* 11 (1): 277–317.

Di Liberto, A., F. Schivardi, and G. Sulis. 2015. "Managerial Practices and Student Performance." *Economic Policy* 30 (84): 683–728.

Donato, K., G. Miller, M. Mohanan, Y. Truskinovsky, and M. Vera-Hernández. 2017. "Personality Traits and Performance Contracts: Evidence from a Field Experiment among Maternity Care Providers in India." *American Economic Review* 107 (5): 506–10.

Fryer, R. G. 2017. "Management and Student Achievement: Evidence from a Randomized Field Experiment." NBER Working Paper W23437, National Bureau of Economic Research, Cambridge, MA.

Karachiwalla, N., and A. Park. 2017. "Promotion Incentives in the Public Sector: Evidence from Chinese Schools." *Journal of Public Economics* 146 (February): 109–28.

King, G., C. J. Murray, J. A. Salomon, and A. Tandon. 2004. "Enhancing the validity and cross-cultural comparability of measurement in survey research." *American Political Science Review* 98(1): 191-207.

Kim, Y., and V. Radoias. 2018. "Screening, Diagnosis, and Long-Term Health Outcomes in Developing Countries—The Case of Hypertension." *PLoS One* 13 (12): e0208466.

Kruk, M. E., G. Nigenda, and F. M. Knaul. 2015. "Redesigning Primary Care to Tackle the Global Epidemic of Noncommunicable Disease." *American Journal of Public Health* 105 (3): 431–37.

Kruk, M. E., M. Paczkowski, G. Mbaruku, H. De Pinho, and S. Galea. 2009. "Women's Preferences for Place of Delivery in Rural Tanzania: A Population-Based Discrete Choice Experiment." *American Journal of Public Health* 99 (9): 1666–72.

Larson, E., J. Sharma, M. A. Bohren, and Ö. Tunçalp. 2019. "When the Patient Is the Expert: Measuring Patient Experience and Satisfaction with Care." *Bulletin of the World Health Organization* 97 (8): 563.

Macarayan, E. K., H. L. Ratcliffe, E. Otupiri, L. R. Hirschhorn, K. Miller, S. R. Lipsitz, A. A. Gawande, and A. Bitton. 2019. "Facility Management Associated with Improved Primary Health Care Outcomes in Ghana." *PLoS One* 14 (7): e0218662.

Min, J., Y. Zhao, L. Slivka, and Y. Wang. 2018. "Double Burden of Diseases Worldwide: Coexistence of Undernutrition and Overnutrition-Related Non-communicable Chronic Diseases." *Obesity Reviews* 19 (1): 49–61.

Rasul, I., and D. Rogger. 2018. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *Economic Journal* 128 (608): 413–46.

Sharma, J., K. Andrews, R. Conner, and R. Gatti. 2020. "What Can the Service Delivery Indicator Surveys Tell Us about COVID-19 Preparedness?" Policy Research Working Paper 9334, World Bank, Washington, DC.

Sharma, J., H. H. Leslie, F. Kundu, and M. E. Kruk. 2017. "Poor Quality for Poor Women? Inequities in the Quality of Antenatal and Delivery Care in Kenya." *PLoS One* 12 (1): e0171236.

Trako, I., E. Molina, and S. Asim. 2019. "The Learning Crisis in Tanzania: Results of the SDI and SABER Service Delivery Surveys." World Bank, Washington, DC.

USAID (United States Agency for International Development). 2014. "Guatemala Nutrition Profile." USAID, Washington DC.

USAID (United States Agency for International Development). 2018. "Madagascar Nutrition Profile." USAID, Washington, DC.

WHO (World Health Organization). 2019. "mhGAP Intervention Guide Version 2.0." WHO, Geneva.

WHO (World Health Organization). 2020. "Continuity of Essential Health Services: Facility Assessment Tool." WHO, Geneva.

World Bank. 2019. "Insights into Disaggregating the Human Capital Index." World Bank, Washington, DC.

World Bank. 2021. *World Development Report 2021: Data for Better Lives.* Washington, DC: World Bank.

World Bank and Government of Kenya. 2019. "Kenya Health Service Delivery Indicator Survey 2018 Report." World Bank, Washington, DC.

Yarrow, N., R. Afkar, E. Masood, and B. Gauthier. 2020. "Measuring the Quality of MoRA's Education Services [English]." World Bank Group, Washington, DC.

5. Lessons from SDI for a post-COVID-19 world

The COVID-19 (coronavirus) pandemic has challenged health and education systems worldwide. Teachers and health care providers have responded to the crisis with redoubled commitment: continuing to work despite increased personal risk and finding creative ways to reach students and care for patients. Health and education systems could not function without the dedication of these frontline workers.

This book has aimed to leverage a decade of Service Delivery Indicators (SDI) data, now fully harmonized across countries, to examine the state of health and education systems, and highlight ways that systems can evolve to help frontline providers to serve patients and students better, during this crisis and beyond. The data presented here were collected in a pre-COVID-19 world, reflecting shortcomings in the quality of service delivery that predate the pandemic. Unfortunately, these shortfalls have lost none of their relevance. If anything, COVID-19 has compounded these systemic weaknesses. This book offers a base of actionable evidence for policy makers as they chart a path forward after COVID-19 and work to make systems stronger and more resilient to future shocks.

Rethinking service delivery

On the health side, the surveys reveal substantial gaps in service quality. Rates of provider absence are high, particularly in public facilities, and the caseloads of staff vary dramatically, with some facilities overwhelmed and others under-attended. Health care providers are correct in only about half of their diagnoses of basic medical conditions, with lower rates among nurses and lower-level health care workers, who are likely to be a patient's first point of contact with the health system. Despite decades of efforts to strengthen the supply chain, equipment and medicines are frequently unavailable. Finally, deficiencies in infrastructure continue to be particularly pronounced in rural areas. The combination of these factors suggests that a typical patient seeking care in these systems is likely to find a facility lacking in the basic necessities for care.

SDI results suggest margins for strengthening the delivery of health services. Health leaders, system coordinators, and facility managers can do more to ensure the presence of health care providers and to balance caseloads fairly by reallocating staff to overburdened facilities. Such efforts may demand politically sensitive trade-offs between widespread geographic presence and improved quality of care. Both presence and caseload can be monitored and actively managed at low marginal cost via improved information management systems. To improve diagnosis, governments will need to reinforce competencies, especially among nonphysician providers in frontline facilities. The lack of equipment and medicine is puzzling given a decades-long focus on increasing the availability of basic inputs. Governments can look at health facilities that succeed on this metric and incentivize other units to emulate these examples. Finally, improving access to water, sanitation, and electricity can go a long way toward reducing infrastructure gaps at rural health facilities.

COVID-19 has brought new challenges to health systems. In addition to the urgency of stopping transmission—with both nonpharmaceutical interventions and vaccinations—policy makers need to protect core functions of health service delivery and ensure equitable access to care, while managing increased stress on the system, including the need for critical care (World Bank 2020a). The pandemic-related recession and growing demands on public expenditures are exerting fiscal pressure on governments. Health spending priorities should be protected in this new environment, including spending for immediate needs, such as providing COVID-19 diagnostics, surveillance, and care, and spending for longer-term objectives, such as expanding universal health coverage. Routine services disrupted by the pandemic will need to resume and, in some cases, address significant lags—for example, in routine immunization. Reckoning with existing system-level weaknesses will be an important step for policy makers and administrators as they embark on the rebuilding process.

In the education sector, SDI evidence gives a granular, frontline picture of the learning crisis so well depicted in the *World Development Report 2018: Learning to Realize Education's Promise* (World Bank 2018). In the nine countries in this sample, measured learning is, on average, low among public school students, but the variance across and within countries is large. Some students in all nine SDI countries are already performing well, illustrating the potential for each country to bring its students to a high level. Results underscore the importance of specific school inputs and teacher characteristics that correlate positively with higher student learning. In particular, classrooms with functioning blackboards; clean, private, and gendered toilets; and more teachers per student are characteristics observed more frequently in schools with higher average student test scores. Equally important is having more knowledgeable teachers, with gender and age being strong predictors of teachers' effectiveness. Finally, private schools in a subsample of SDI countries show higher average levels of learning than their public counterparts.

In already challenging contexts, governments should not take for granted basic inputs that can help to improve students' daily experience in school. Having functioning blackboards and toilets might seem trivial, but evidence suggests that these features can make a big difference for some children. The SDI evidence suggests that policy makers should prioritize hiring, retaining, and continuously training more and better teachers to improve student learning. Content knowledge and pedagogical skills, two strong predictors of learning, should be incorporated in hiring practices and professional development systems for teachers. Additional lessons and insights may arise from studying education systems as a whole, including the private sector, which seems to be doing better, on average, in a subset of countries. Education SDI surveys can help governments to deploy investments efficiently so that more young people can reach their full potential.

Today, COVID-19 threatens educational outcomes in many countries (World Bank 2020b). The combination of school closures and economic recession is likely to increase dropout rates and affect learning significantly. Disconcertingly, it is often the most vulnerable members of the population who lose access to education. Strategies to remediate schooling losses will require designing and implementing school reopening protocols adapted to the specificities of the pandemic. At a minimum, these efforts will involve protective equipment and supplies, health screening, and social distancing. Tailored teaching and learning resources, especially for disadvantaged children, are urgently needed in many settings to make up for lost learning (World Bank 2020c). Deeper reforms will be needed to sustain access to schooling and promote children's learning at all stages, from cognitive stimulation in the early years to nurturing relevant skills in childhood and adolescence. Building blocks for success will include

better-prepared teachers, better-managed schools, and incentives that are aligned across the many stakeholders in education reform.

Although it will take time to stabilize education systems, this book has summarized some of the areas most in need of rebuilding and rehabilitation. High-quality primary education is one of the best investments that countries can make to drive productivity gains and economic growth. Ensuring student learning across the whole population should remain a high priority as governments rebuild after COVID-19.

Moving forward, digital technology will play an even more important role in both education and health. The efforts that countries have made in providing continuity with remote learning during the pandemic could carry benefits beyond the current emergency. Appropriately structured online learning can facilitate the acquisition of competencies essential in the changing world of work, such as collaboration and higher-order cognitive skills (Reimers and Schleicher 2020). To shape resilient education systems, countries will need to draw lessons from worldwide experience with distance learning and expand the infrastructure for online and remote learning. In health, telemedicine can, in some cases, help to increase access to higher-quality care in an efficient manner. Finally, in the process of moving toward digital solutions, governments should consider specific disadvantages that some groups face in accessing online platforms, such as gender and disability gaps in access to and use of technology.

SDI surveys: Turning measurement into momentum for reform

Measurement is essential to motivate and guide reforms in health and education, the core pillars of human capital. SDI surveys provide clear, actionable indicators that highlight where health and education systems function well and where fixes are needed. The surveys can help to prioritize reform steps and track the effects of reforms in health and education over time. They also pinpoint areas of reform in governance and accountability that apply to both sectors.

Perhaps most important, SDI results can motivate change by offering an open and transparent assessment of system performance that policy makers and citizens alike can understand. Although reforms often come from the top down, the SDI surveys are premised on the conviction that change can be driven from the bottom up, by engaging citizens to demand high-quality services. The results presented here can function as a call for change in many countries.

As evidenced in the recent *World Development Report 2021: Data for Better Lives*, the commitment to measurement among policy makers and stakeholders continues to gain force globally (World Bank 2021). The SDI surveys are part of that agenda. Over the next decade, health and education management information systems will likely expand and incorporate a greater range and quality of data, including electronic medical records, student performance data, and more advanced monitoring of facilities. However, the deployment and refinement of these systems will take time, particularly in countries recovering from the pandemic and the ensuing economic crisis. SDI health and education surveys will continue to play an important role by allowing greater breadth of indicators, not as substitutes for but as complements to administrative data. Efforts to ensure compatibility between these two sources will require that surveys be flexible and well adapted to country needs, while also retaining their rigor and international comparability. The performance of health and education systems can be measured in multiple ways, and, by looking at data from all angles and using diverse instruments, policy makers will be able to see the strengths and weaknesses of systems more clearly.

The SDI surveys have expanded geographically in recent years, reflecting the global demand for reliable measurement. SDI efforts in Afghanistan, Armenia, Bhutan, Cameroon, the Comoros, Côte d'Ivoire, the Democratic Republic of Congo, Guatemala, Guinea-Bissau, Indonesia, Iraq, Malawi, Moldova, and Pakistan are ongoing or were recently completed during the production of this book. The health surveys in Bhutan, Guatemala, and Moldova have required substantial adaptation and innovation to suit the new contexts. This adaptation has meant an expanded focus, with methodological innovations to measure more health conditions, including nutrition, mental health, and noncommunicable diseases. The surveys have also become more people-centered, focusing on the experience of providers and patients. Likewise, recent SDI education surveys in the Democratic Republic of Congo and Indonesia have brought innovations needed to measure additional aspects of education, such as teachers' and students' socioemotional skills, children's perceptions of their teachers, parents' school choice and socioemotional conditions, and households' socioeconomic characteristics. The expanded geography of the SDI surveys will bring both new learning and new challenges, requiring greater flexibility in the survey tool but also opening new areas for innovation.

As COVID-19 vaccines slowly become globally available, students will return to school, and doctors and nurses will resume routine practice. This progressive return to normality will be comforting in some respects. But the crisis has highlighted many fault lines in what was previously considered normal. This crisis has provided an opportunity to reinvigorate and transform systems.

This book has shown that health and education systems in many countries are falling far short of what they could achieve—and what citizens expect. These systems deliver inadequate and uneven services, hampering the human capital development on which communities' well-being and countries' economic progress depend. Countries' efforts to end poverty and advance shared prosperity by reigniting inclusive, sustainable growth can succeed only if all people are able to access quality health care and education. As they evolve to meet policy makers' needs and reflect citizens' experience with systems, the SDI surveys will provide measurement and accountability to support this agenda.

References

Reimers, F., and A. Schleicher. 2020. "Schooling Disrupted, Schooling Rethought: How the Covid-19 Pandemic Is Changing Education." Organisation for Economic Co-operation and Development, Paris.

World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise.* Washington, DC: World Bank.

World Bank. 2020a. "COVID-19 and Health Systems: Just the Beginning." World Bank, Washington, DC.

World Bank. 2020b. "The COVID-19 Pandemic: Shocks to Education and Policy Responses." World Bank, Washington, DC.

World Bank. 2020c. "From Data to Action: Big Data, GIS Tools, and Machine Learning to Modernize the Delivery of Health Services in Peru." World Bank, Washington, DC.

World Bank. 2021. *World Development Report 2021: Data for Better Lives.* Washington, DC: World Bank.

Appendixes

Appendix A: Supplemental tables

TABLE A.1   Sample of SDI health surveys, disaggregated by country, location, facility level, and ownership

Indicator	All countries		Niger (2015)		Mozambique (2014)		Sierra Leone (2018)		Madagascar (2016)		Togo (2013)		Uganda (2013)		Tanzania (2016)		Kenya (2018)		Nigeria (2013)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
GDP per capita in survey year (US$)			1,048		1,328		1,604		1,634		1,761		2,033		3,227		3,461		5,980	
Total facilities	7,810	100	255	100	195	100	536	100	444	100	180	100	394	100	383	100	3,038	100	2,385	100
Rural	5,276	68	192	75	172	88	377	70	218	49	126	70	285	72	222	58	2,249	74	1,435	60
Urban	2,534	32	63	25	23	12	159	30	226	51	54	30	109	28	161	42	789	26	950	40
Hospitals	872	11	16	6	38	19	30	6	37	8	16	9	9	2	30	8	285	9	411	17
Health clinics	2,823	36	67	26	157	81	109	20	316	71	46	26	133	34	91	24	594	20	1,458	61
Health posts	4,115	53	172	67	0	0	397	74	91	20	118	66	252	64	262	68	2,159	71	516	22
Public	5,805	74	220	86	193	99	493	92	289	65	143	79	236	60	266	69	1,762	58	2,203	92
Private or NGO	2,005	26	35	14	2	1	43	8	155	35	37	21	158	40	117	31	1,276	42	182	8

Source: Service Delivery Indicators (SDI) health surveys.
Note: The survey in Nigeria included 12 of 36 states because of logistical constraints, and the survey in Kenya was representative at the national and county levels.
NGO = nongovernmental organization.

TABLE A.2 Sample of SDI education surveys, disaggregated by location, ownership, and student body gender

Indicator	All countries		Kenya (2012)		Madagascar (2016)		Mozambique (2014)		Niger (2015)		Nigeria (2013)		Tanzania (2016)		Togo (2013)		Uganda (2013)		Morocco (2016)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
GDP per capita in survey year (US$)			3,461		1,634		1,328		1,048		5,980		3,227		1,761		2,033		8,587	
Total schools	3,297	100	306	100	473	1	100	1	256	100	760	100	400	100	200	100	400	100	299	100
Rural	2,406	73	207	68	319	67	172	85	169	66	604	79	314	79	136	68	325	81	160	54
Urban	891	27	99	32	154	33	31	15	87	34	156	21	86	22	64	32	75	19	139	46
Public	2,537	77	239	78	308	65	203	100	223	87	458	60	397	99	147	74	319	80	243	81
Private	758	23	67	22	165	35	0	0	33	13	300	39	3	1	53	27	81	20	56	19
Boys	57	2	10	3	0	0	0	0	3	1	27	4	1	0	0	0	16	4	0	0
Girls	40	1	7	2	0	0	0	0	1	0	21	3	6	2	1	1	4	1	0	0
Coed	3,199	97	289	94	473	100	203	100	252	98	712	94	393	98	198	99	380	95	299	100

Source: Service Delivery Indicators (SDI) education surveys.
Note: The survey in Nigeria included four states: Anambra, Bauchi, Ekiti, and Niger. For the Kenya survey, three counties of North Eastern Province were excluded because of security concerns.

TABLE A.3 Definition and calculation of health indicators

Indicator	Definition	Details
Infrastructure availability	Availability of an improved water source, an improved toilet, and electricity	Credit is given if all three components are available. *Improved toilet:* Credit is given if facility reports and enumerator confirms that the facility has one or more functioning flush toilet, ventilated improved pit latrine, or covered pit latrine (with slab). *Improved water source:* Credit is given if facility reports that its main source of water is piped into the facility, piped onto the facility's grounds, or comes from a public tap or standpipe, tube well or borehole, a protected dug well, a protected spring, bottled water, or a tanker truck. This definition is based on the World Health Organization (WHO) and United Nations Children's Fund (UNICEF) Joint Monitoring Program for Water Supply, Sanitation, and Hygiene (WHO and UNICEF 2019).
Medicine availability	Percent of 14 basic medicines that are available and in stock at the time of the survey	The SDI list of medicines is based on a subset of the WHO list of essential medicines (WHO 2014). The following medicines are included: (1) amitriptyline (antidepressant), (2) amoxicillin (antibiotic), (3) atenolol (beta blocker), (4) captopril (ACE inhibitor), (5) ceftriaxone (antibiotic), (6) ciprofloxacin (antibiotic), (7) cotrimoxazole (antibiotic), (8) diazepam (antiseizure), (9) diclofenac (nonsteroidal anti-inflammatory) (10) glibenclamide (antidiabetic), (11) omeprazole (proton pump inhibitor), (12) paracetamol (analgesic), (13) salbutamol (bronchodilator), and (14) simvastatin (statin). The SDI list of medicines is adapted to each country's standards; some of these medicines were not included in the surveys in Kenya, Nigeria, and Uganda, so these countries were omitted from this indicator.
Equipment availability	Availability and functioning of a thermometer, stethoscope, sphygmomanometer, and weighing scale (adult, child, infant)	Credit is given if all four components are available. *Thermometer:* Credit is given if the facility reports and the enumerator observes that the facility has one or more functioning thermometer (used for measuring body temperature). *Stethoscope:* Credit is given if the facility reports and the enumerator observes that the facility has one or more functioning stethoscope. *Sphygmomanometer:* Credit is given if the facility reports and the enumerator observes that the facility has one or more functioning sphygmomanometer. *Weighing scale:* Credit is given if the facility reports and the enumerator observes that the facility has one or more functioning adult, child, or infant weighing scale.
Caseload per health care provider	Number of outpatient visits per clinician per day	Caseload is calculated as the number of outpatient visits recorded in outpatient records in the three months prior to the survey, divided by the number of days the facility was open during the three-month period and the number of health professionals who conduct patient consultations. This indicator is adjusted for the average absenteeism at the facility level. For example, if a facility reports having 10 health care providers who conduct outpatient consultations, but its absenteeism on an unannounced visit is found to be 40%, then the number of health care providers is adjusted down by 40% and only 6 health care providers are counted as available for patient care.
Provider absenteeism	Share of randomly selected providers absent from the facility during an unannounced visit	Number of health professionals who are absent from the facility on an unannounced visit as a % of randomly sampled workers who should be on duty. Health professionals doing outreach are counted as present.
Diagnostic accuracy	Share of correct diagnoses provided in the five clinical vignettes	The SDI includes five core vignettes: (1) acute diarrhea with dehydration, (2) pneumonia, (3) diabetes mellitus, (4) pulmonary tuberculosis, and (5) malaria with anemia. Health care providers are scored on their ability to provide a correct diagnosis on each of the vignettes, and their overall score is calculated as the share of vignettes answered correctly. Further details on diagnostic accuracy for each disease are provided in table A.5.
Treatment accuracy	Share of correct treatments provided in the five clinical vignettes	The SDI includes five core vignettes: (1) acute diarrhea with dehydration, (2) pneumonia, (3) diabetes mellitus, (4) pulmonary tuberculosis, and (5) malaria with anemia. Health care providers are scored on their ability to provide correct treatment on each of those vignettes, and their overall score is calculated as the % of vignettes answered correctly. Further details on treatment accuracy for each disease are provided in table A.6.

(continued)

TABLE A.3   Definition and calculation of health indicators (*Continued*)

Indicator	Definition	Details
Management of maternal and neonatal complications	Number of relevant treatment actions proposed by the clinician	The SDI includes two vignettes to assess maternal and neonatal complications. Providers are scored on the number of relevant treatment actions that they propose out of five specific actions for postpartum hemorrhage and seven specific actions for neonatal asphyxia.
Inappropriate antibiotic use	Share of providers inappropriately prescribing antibiotics	The SDI surveys also collect information on inappropriate antibiotic use, defined as prescribing an antibiotic during the tuberculosis vignettes (aside from the antibiotics recommended as part of the tuberculosis regimen) or any antibiotics for the diarrhea vignette (for which antibiotics are not indicated by the patient examination). Inappropriate antibiotic use is calculated as the % of health care providers who inappropriately prescribe antibiotics among all health care providers given the clinical vignettes.

Source: Service Delivery Indicators (SDI) core team.

TABLE A.4   Definition of education indicators

Indicator	Definition	Details
School absence rate	Share of a maximum of 10 randomly selected teachers who are absent from school during an unannounced visit	This indicator is measured in the following way: During the first, announced, visit, a maximum of 10 teachers are randomly selected from the list of all teachers (excluding volunteer and part-time teachers) on the school roster. The whereabouts of these 10 teachers are then verified during the second, unannounced, visit. Teachers found anywhere on the school premises are marked as present.
Classroom absence rate	Share of teachers who are present in the classroom during scheduled teaching hours as observed during an unannounced visit	This indicator is constructed in the same way as the school absence rate indicator, except that the numerator is the number of teachers who are either absent from school or present at school but absent from the classroom.
Time spent teaching per day	Amount of time a teacher spends teaching during an average school day	This indicator combines data from the staff roster module (used to measure absence rate), the classroom observation module, and reported teaching hours. Teaching time is adjusted for the time teachers are absent from the classroom, on average, and for the time teachers teach while in the classroom, based on classroom observations. While inside the classroom, distinction is made between teaching and nonteaching activities. Teaching is defined very broadly, including actively interacting with students, correcting or grading students' work, asking questions, testing, using the blackboard, or having students work on a specific task, drill, or memorization. Nonteaching activities include working on private matters, maintaining discipline in class, or doing nothing and thus having students who are not paying attention.
Minimum knowledge	Share of teachers with minimum knowledge and test score	This indicator is based on mathematics and language tests covering the primary curriculum administered at the school and is calculated as the % of teachers who score more than 80% on the language and mathematics portion of the test. The test is given to all mathematics or language teachers who taught third grade in the prior year or fourth grade in the year the survey is conducted. Test score is measured as the overall score on mathematics, language, and pedagogy tests covering the primary curriculum administered at the school level to all mathematics and language teachers who taught third grade in the prior year or fourth grade in the year the survey is conducted.

(continued)

TABLE A.4 Definition of education indicators (*Continued*)

Indicator	Definition	Details
Minimum infrastructure availability	Unweighted average of the proportion of schools with the following available: functioning electricity and sanitation	This is a binary indicator capturing availability of (1) functioning toilets and (2) classroom visibility. Functioning toilets is defined as whether toilets are functioning, accessible, clean, and private (enclosed and with gender separation), as verified by an enumerator. To verify classroom visibility, the survey team randomly selects one fourth-grade classroom in which the enumerator places a printout on the board and checks whether it is possible to read the printout from the back of the classroom.
Minimum equipment availability	Unweighted average of the proportion of schools with the following available: functioning blackboard with chalk, pens or pencils, and notebooks or paper	This is a binary indicator capturing availability of (1) functioning blackboard and chalk and (2) pens, pencils, and exercise books in fourth-grade classrooms. In one randomly selected fourth-grade classroom in the school, the enumerator assesses whether there is a functioning blackboard by looking at whether text written on the blackboard can be read at the front and back of the classroom and whether chalk is available to write on the blackboard. A classroom has met the minimum requirement of pens, pencils, and exercise books if both the share of students with pen or pencils and the share of students with exercise books are above 90%.
Share of pupils with textbooks	Number of mathematics and language books used in a fourth-grade classroom divided by the number of pupils present in the classroom	This indicator reflects the typical ratio of students to textbooks in the fourth-grade classroom. It is measured as the number of students with the relevant textbooks (mathematics or language, conditional on which class is observed) in one randomly selected fourth-grade class divided by the number of students in that classroom.
Observed pupil-teacher ratio	Average number of fourth-grade pupils per fourth-grade teacher	This indicator reflects the typical ratio of pupils to teacher in the fourth-grade classroom. It is measured as the number of students in one randomly selected fourth-grade class at the school.

Source: Service Delivery Indicators (SDI) core team.

TABLE A.5 Definition of a correct diagnosis

Disease	Notes
Diarrhea with dehydration	The integrated management of childhood illness (IMCI) guidelines (WHO 2014) suggest that this case should be classified as diarrhea with severe dehydration due to the presence of three warning signs: lethargy, sunken eyes, and skin pinch going back very slowly. Clinicians should arrive at the dual diagnosis of diarrhea and dehydration but are allowed a broader range of classifications than suggested by the IMCI guidelines. These diagnoses include any mention of diarrhea ("diarrhea" or "acute diarrhea") and any mention of dehydration ("dehydration," "moderate dehydration," or "severe dehydration"). In Mozambique and Niger, the child was listed as unable to drink or drinking poorly, another risk sign for severe dehydration. The way that the diagnosis for diarrhea with dehydration was recorded varied slightly from country-to-country. For example, Nigeria and Uganda listed "acute diarrhea with severe dehydration" as the only possible diagnosis. Madagascar listed "diarrhea with moderate dehydration" and "diarrhea with severe dehydration." Togo listed "diarrhea with severe dehydration." All of these answers are counted as correct. For these countries, it is not possible to provide an accurate estimate of how many doctors diagnosed diarrhea alone; it is only possible to calculate the joint diagnosis rate.
Pneumonia	A diagnosis of pneumonia is counted as the correct response.
Diabetes	Based on the characteristics presented in this vignette, a diagnosis of type 2 diabetes is the correct response. However, the option of diabetes (type not specified) was available in Kenya, Madagascar, Mozambique, Niger, Sierra Leone, and Tanzania. Clinicians were not prompted to select a specific type if they answered diabetes. Although treatment varies for different types of diabetes, the general diabetes response is also classified as correct. This practice results in higher correct diagnosis rates, an improvement by 14–50 percentage points depending on the country.
Tuberculosis	A diagnosis of tuberculosis is counted as the correct response.
Malaria with anemia	This case should be classified as malaria with anemia, and clinicians are required to arrive at this dual diagnosis. All countries except Togo included "malaria" as an option, and all countries except Nigeria and Uganda included "simple malaria" as an option. Both of these diagnoses are counted as correct. All countries also included "severe malaria" as an option, and this nonspecific diagnosis is not counted as correct because the case does not meet the definition of severe, and severe malaria would require different treatment. Anemia was listed simply as "anemia" without specifying severity. Providers therefore received credit for specifying malaria or simple malaria and anemia. Kenya excluded this module entirely, so it is omitted and its diagnostic accuracy is counted as the average of the four other vignettes.
Postpartum hemorrhage	A diagnosis of postpartum hemorrhage is counted as the correct response.
Neonatal asphyxia	A diagnosis of neonatal asphyxia or respiratory distress is counted as the correct response.

Source: Service Delivery Indicators (SDI) core team.

TABLE A.6 Definition of a correct treatment

Disease	Notes
Diarrhea with dehydration	The World Health Organization (WHO) guidelines on the integrated management of childhood illness (IMCI) note that the correct treatment of diarrhea with severe dehydration is to give intravenous fluid immediately, to insert a nasogastric (NG) tube if that is not possible, and to refer the patient to a higher-level facility if neither treatment is available (WHO 2019b). If the child has only some dehydration, then oral rehydration salts (ORS) are the recommended treatment. Given the symptoms, the correct treatment should be rehydration with an intravenous (IV) line or an NG tube. However, the use of ORS plus zinc is also counted as correct. Because the child was able to drink in most vignettes, the providers may have incorrectly believed that the dehydration was less severe. Correct treatment rates would be much lower if only IV fluids or an NG tube was counted as correct (13% correct). Uganda did not include an option for ORS, so only treatment with IV fluids or an NG tube is taken as a correct response. Kenya did not include an option for IV fluids or NG tube, so only ORS with zinc is taken as a correct response.
Pneumonia	The IMCI guidelines suggest oral amoxicillin for five days as treatment for pneumonia. Severe pneumonia can be treated with "the first dose of an appropriate antibiotic" and urgent referral to a hospital. In addition, children had a fever of 38.5 in the vignette, and IMCI guidelines recommend an antipyretic in this case. Correct treatment is counted as treatment with amoxicillin, a first-line antibiotic, and any antipyretic.
	The WHO package of essential noncommunicable disease interventions (PEN) protocols (WHO 2019d) states, "Individuals with persistent fasting blood glucose >6 mmol/l despite diet control should be given metformin and/or insulin as appropriate." Correct treatment is counted as any hypoglycemic (including insulin) or referral to a specialist. Referral to a higher level is the recommended protocol for diabetes at the primary level in multiple countries so that option is counted as correct. Although PEN protocol suggests diet control before prescribing hypoglycemics, here prescriptions on first presentation are counted as correct.
Tuberculosis	The WHO guidelines for treatment of tuberculosis (WHO 2018) recommend combination therapy, ideally with a fixed-dose combination. Providers are simply required to mention combination therapy. Knowledge of correct duration and dosage is not necessary, and providers would score worse if this knowledge were required. For example, 23% of providers prescribed combination therapy, but only 8% accurately recalled the correct dosage and timing (this comparison is possible in Madagascar, Mozambique, Niger, Sierra Leone, and Tanzania). However, Nigeria and Uganda recorded "correct duration and dose" as one option, so providers are assessed on having gotten the correct dosage and timing. This likely creates a downward bias for provider treatment abilities in these two countries. Kenya did not record any information on whether providers got the correct duration and dosage.
Malaria with anemia	IMCI guidelines recommend that children with a positive malaria test should be given "recommended first-line antimalarial" and "one dose of paracetamol in clinic" for fever reduction. In addition, iron should be given for treatment of anemia. The questions on malaria treatment varied a bit between countries, and credit is given for treatment with any artemisinin combination therapy or artemether-lumefantrin (coartem). In addition to antimalarials, the provider must prescribe paracetamol and iron for the anemia. Kenya did not include the malaria vignette and is excluded. Nigeria and Uganda did not include questions about iron and are excluded for the sake of comparability.
Postpartum hemorrhage (PPH)	The WHO recommendations for the prevention and treatment of postpartum hemorrhage (WHO 2019c) state, "The use of uterotonics (oxytocin alone as the first choice) plays a central role in the treatment of PPH. Uterine massage is recommended for the treatment of PPH as soon as it is diagnosed, and initial fluid resuscitation with isotonic crystalloids is recommended." A provider should propose five specific actions: (1) determining a cause, (2) prescribing any uterotonics, (3) inserting an IV line, (4) inserting a foley catheter, (5) prescribing bimanual uterine massage. Uterotonics include oxytocin or another uterotonic drug, such as misoprostol, ergometrine, or a prostaglandin. In Togo and Uganda, health care providers are given credit only if they were able to determine the correct dosage of oxytocin.
Neonatal asphyxia	The WHO guidelines on basic newborn resuscitation (WHO 2019a) provide detailed recommendations on newborn care and actions that should be taken if a child is exhibiting danger signs. Assessments for neonatal asphyxia varied between countries. Seven actions are assessed in all countries, and these actions are used as the SDI standard of assessment: (1) call for help, (2) dry the baby, (3) keep the baby warm, (4) check to see if the baby is breathing, (5) place the baby in a natural position, (6) initiate resuscitation with a bag or mask, and (7) check the baby's heart rate. This set of actions does not represent the full standard of treatment for neonatal asphyxia but does allow for a common set of actions on which health care providers can be compared across the sample of countries. Correct treatment is counted as providing at least half of these actions.

Source: Service Delivery Indicators (SDI) core team.

TABLE A.7 How do inputs differ across high- and low-performing schools?

Variable	(1) Bottom		(2) Top		(3) All		t-test Difference
	N	Mean	N	Mean	N	Mean	(1)–(2)
Functioning toilets (available + clean + private + gendered)	127	0.150 [0.032]	126	0.468 [0.045]	2,537	0.381 [0.010]	−0.319***
Visibility indicator (enumerator, front of the classroom)	127	0.858 [0.031]	125	0.920 [0.024]	2,513	0.888 [0.006]	−0.062
Share of observed students with pen or pencils	127	0.811 [0.026]	124	0.938 [0.019]	2,510	0.933 [0.004]	−0.127***
Share of observed students with exercise book	127	0.752 [0.031]	126	0.910 [0.024]	2,518	0.884 [0.005]	−0.159***
Share of observed students with textbook	127	0.227 [0.031]	126	0.427 [0.038]	2,509	0.356 [0.008]	−0.200***
Board functioning indicator	126	0.730 [0.040]	126	0.937 [0.022]	2,531	0.829 [0.007]	−0.206***
Multigrade classes indicator	127	0.291 [0.040]	126	0.175 [0.034]	2,531	0.213 [0.008]	0.117**
Pupil-teacher ratio	127	56.055 [3.827]	126	34.683 [1.787]	2,535	40.280 [0.486]	21.371***

Source: Service Delivery Indicators (SDI) education surveys.
Note: The values displayed for *t*-tests are the differences in the means across the groups. Numbers in brackets are standard errors.
** p<0.05   *** p<0.01

TABLE A.8 How do teacher traits differ across high- and low- performing schools?

Variable	(1) Bottom		(2) Top		(3) All		t-test Difference
	N	Mean	N	Mean	N	Mean	(1)–(2)
Mean teacher test score (task 2 only) [IRT MLE]	126	−1.403 [0.134]	126	−0.019 [0.091]	2,274	−0.322 [0.026]	−1.384***
Mean teacher test score (task 1 only) [IRT MLE]	126	−0.850 [0.120]	126	0.396 [0.125]	2,274	0.076 [0.026]	−1.246***
Mean teacher test score jointly estimated across tasks 3+4+5 [IRT EAP]	126	−0.745 [0.058]	123	−0.101 [0.064]	2,265	−0.188 [0.013]	−0.644***
Mean teacher age	127	35.323 [0.616]	126	39.524 [0.671]	2,280	37.710 [0.133]	−4.201***
Proportion of male teachers	127	0.797 [0.021]	126	0.402 [0.024]	2,281	0.529 [0.006]	0.396***
Proportion of teachers with less than primary education	127	0.265 [0.035]	126	0.126 [0.022]	2,281	0.194 [0.006]	0.138***
Proportion of teachers with secondary education	127	0.290 [0.028]	126	0.350 [0.028]	2,281	0.373 [0.007]	−0.060

(continued)

TABLE A.8 How do teacher traits differ across high- and low- performing schools? (*Continued*)

Variable	(1) Bottom		(2) Top		(3) All		t-test Difference
	N	Mean	N	Mean	N	Mean	(1)–(2)
Proportion of teachers with above secondary education	127	0.365 [0.037]	126	0.322 [0.031]	2,281	0.294 [0.007]	0.042
Proportion of teachers with any certification	126	0.820 [0.023]	126	0.652 [0.031]	2,278	0.731 [0.006]	0.168***
Share of time teaching	126	0.810 [0.029]	126	0.751 [0.024]	2,278	0.785 [0.005]	0.059
Mean teacher absence on second visit either from the class or the school (unweighted)	125	0.352 [0.031]	123	0.286 [0.025]	2,231	0.360 [0.006]	0.066

Source: Service Delivery Indicators (SDI) education surveys.
Note: The values displayed for *t*-tests are the differences in the means across the groups. Numbers in brackets are standard errors. IRT MLE=item response theory with maximum likelihood estimator. IRT EAP= item response theory with expected a posteriori estimator.
*** p<0.01

TABLE A.9 Explaining variation in student test scores

Indicator	(1)	(2)	(3)	(4)	(5)
Board functioning indicator	0.179***	0.174***	0.148***	0.151***	0.142**
	(0.050)	(0.050)	(0.050)	(0.050)	(0.060)
Toilets available + private + gendered		0.087*	0.073	0.073	0.060
		(0.045)	(0.045)	(0.045)	(0.052)
Pupil-teacher ratio			−0.006***	−0.006***	−0.005***
			(0.001)	(0.001)	(0.001)
Teacher characteristics principal components analysis	0.231***	0.227***	0.208***	0.200***	0.221***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.030)
Dummy = 1 if school is in urban area	0.243***	0.233***	0.217***	0.155***	0.111**
	(0.039)	(0.039)	(0.039)	(0.044)	(0.055)
Travel time category: 30–60 minutes				−0.213**	−0.162*
				(0.088)	(0.096)
Travel time category: 1–2 hours				−0.268***	−0.203**
				(0.076)	(0.095)

(continued)

TABLE A.9 Explaining variation in student test scores (*Continued*)

Indicator	(1)	(2)	(3)	(4)	(5)
Travel time category: 2–4 hours				−0.257***	−0.182**
				(0.074)	(0.090)
Travel time category: 4+ hours				−0.225***	−0.126
				(0.073)	(0.094)
Poverty proxy (ln night lights)					0.013
					(0.010)
Constant	−0.507***	−0.548***	−0.283***	−0.069	−0.126
	(0.045)	(0.049)	(0.066)	(0.087)	(0.096)
Adjusted R^2	0.390	0.391	0.403	0.406	0.408
Observations	2,453	2,453	2,453	2,453	1,979

Source: Service Delivery Indicators (SDI) education surveys.
Note: Country fixed effects included in all specifications, and robust errors are shown.
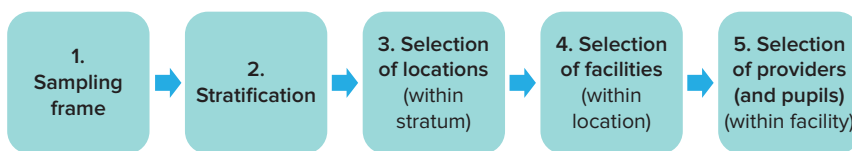* $p<0.10$ ** $p<0.05$ *** $p<0.01$

References

WHO (World Health Organization). 2014. "IMCI Chart Booklet." WHO, Geneva, March 2014. https://www.who.int/docs/default-source/mca-documents/imci-chart-booklet. pdf?sfvrsn=f63af425_1&download=true.

WHO (World Health Organization). 2018. "Guidelines for Treatment of Drug-Susceptible Tuberculosis and Patient Care (2017 Update)." WHO, Geneva, May 18, 2018. www.who.int/tb/publications/2017/dstb_guidance_2017/en/.

WHO (World Health Organization). 2019a. "Guidelines on Basic Newborn Resuscitation: Maternal, Newborn, Child, and Adolescent Health." WHO, Geneva, February 8, 2019. www.who.int/maternal_child_adolescent/documents/basic_newborn_resuscitation/en/.

WHO (World Health Organization). 2019b. "Model List of Essential Medicines." Geneva: WHO.

WHO (World Health Organization). 2019c. "Recommendations for the Prevention and Treatment of Postpartum Haemorrhage: Sexual and Reproductive Health." WHO, Geneva, February 6, 2019. www.who.int/reproductivehealth/publications/maternal_perinatal_health/9789241548502/en/.

WHO (World Health Organization). 2019d. "WHO Package of Essential Noncommunicable (PEN) Disease Interventions: Noncommunicable Diseases and Their Risk Factors." WHO, Geneva. July 22, 2019, www.who.int/ncds/management/pen_tools/en/.

WHO (World Health Organization) and United Nations Children's Fund (UNICEF). 2019. *Water, Sanitation, and Hygiene in Health Care Facilities: Global Baseline Report 2019.* Geneva: WHO and UNICEF.

Appendix B: Sampling procedures

The Service Delivery Indicators (SDI) surveys are designed to attain indicators that are accurate and representative at the national level, allowing for proper cross-country benchmarking and comparisons over time, when applicable.

In addition, other levels of representativeness are sought to allow for further disaggregation during the analysis. Often, samples are constructed to be representative of subregions (regions or provinces), rural and urban areas, public and private providers, and other relevant features (for example, type of facility, performance, or others). Disaggregation requires some level of stratification,[1] which improves the precision of indicators and allows for more accurate within-country comparisons to provide more targeted and granular policy recommendations. It might also require oversampling in certain geographic locations to capture data adequately. Power calculations using presurvey data are computed to estimate required sample sizes and minimum detectable effect sizes.

Essentially, the sampling strategy for SDI surveys follows a multistage sampling approach. The main units of analysis are facilities (schools and health centers) and providers (health and education workers: teachers, doctors, nurses, and others). In the case of education, SDI surveys also aim to produce accurate information (that is, within a desired confidence interval) on the performance

FIGURE B.1 Typical sampling strategy process for SDI surveys

| 1. Sampling frame | → | 2. Stratification | → | 3. Selection of locations (within stratum) | → | 4. Selection of facilities (within location) | → | 5. Selection of providers (and pupils) (within facility) |

Source: Service Delivery Indicators (SDI) core team.

of fourth-grade pupils through a student assessment. The multistage sampling approach makes sampling procedures more practical by dividing the selection of large populations of sampling units in a step-by-step fashion (figure B.1). After defining the sampling frame (that is, the complete list from which sampling units are drawn)[2] and categorizing it by stratum, a first-stage selection of sampling units is carried out independently within each stratum. Often, the first stage is selecting clusters or geographic locations (districts, communities, counties, neighborhoods) to ensure that survey teams do not have to travel long distances to interview just one facility. Clusters are randomly drawn within each stratum with a probability proportional to the size of the cluster (measured by the location's number of facilities, providers, or pupils), which helps to ensure that the sample is representative of the services within those locations. Once locations are selected, a second stage takes place by randomly selecting facilities within locations (either with equal probability or with probability proportional to size) as secondary sampling units.[3] At a third stage, a fixed number of health and education workers and pupils is randomly selected within facilities to provide information for the different questionnaire modules.

Replacement facilities are also drawn from each location strata in case the sampling frame includes health or school facilities that no longer exist, are not functional, refuse to participate, or are inaccessible because of security concerns. These replacement facilities are selected in keeping with the probability sampling approach. More important, the rules for replacement are specified in the protocol ex ante to avoid bias in the results, and backup facilities are typically not allowed to be used for logistical ease.[4]

Because of this sampling process, survey results must be properly weighted using a sampling weight or expansion factor to assure representativeness of the population of interest. The basic weight for each sampling unit is equal to the inverse of its overall probability of selection, which is computed by multiplying the probabilities of selection at each sampling stage. Different weights need to be applied depending on the relevant level for the estimate, which can be the facility, the staff or provider, or the pupil.[5] These different weights are later included in the data sets to facilitate reestimations.

Sampling procedures must also adapt to country-specific contexts. In a few countries, certain types of facilities represent a very small fraction of the entire roster of facilities and, thus, can be excluded from the sampling frame. For example, private rural schools were not included in the SDI education surveys in Niger (2015). Likewise, in Mozambique (2014) and Tanzania (2016), private schools represented 3 percent and 1 percent of the schooling system, respectively, and were therefore excluded for practical reasons. For the 2013 SDI health survey in Nigeria, tertiary facilities were excluded from the surveys, because there were very few of them (11 facilities across six states), and they would not be surveyed appropriately with the same instrument used for primary or secondary facilities.

Similarly, other facilities are sometimes discarded because they are not eligible for the SDI. For example, community schools and special-needs schools not formally recognized by the ministry of education as well as "American" or "English" schools (which adhere to curricula distinct from the national curriculum) were not included in the 2013 SDI education surveys in Togo. For the 2015 SDI education survey in Niger, *écoles franco arabes* were not considered (they followed a Koranic curriculum). For the 2014 SDI health survey in Tanzania, 93 facilities (approximately 1.2 percent of the total universe of health facilities) were deleted from the frame because they correspond to regional hospitals, dental clinics, specialized clinics, or facilities that serve prison populations.

Likewise, locations can be discarded from the sampling process because of security concerns (for example, areas that are under a state of emergency or undergoing active conflict) or other logistical restrictions. In Kenya, three counties of North Eastern Province were excluded from the 2012 SDI education survey because of security concerns. Similarly, in Niger in 2015, schools from the region of Diffa, which was in a state of emergency, were excluded from the SDI education surveys. For the 2014 SDI health surveys in Mozambique, the sample was reduced from 300 health facilities originally to 204 facilities because of logistical and financial problems. However, these exclusions did not affect the representativeness of results at the national level.

There have been very few exceptions when, because of budgetary constraints, SDI surveys could not achieve national representativeness. For instance, in Nigeria (2013), the SDIs on education and health were representative of 4 and 12 states respectively, mostly because of the size of the country. The surveys are, nonetheless, representative of each of the selected states.

Finally, some locations or facilities are selected because they are deemed unique in the sample or their characteristics represent opportunities for interesting comparisons or analysis. For instance, for the 2012 SDI education survey in Kenya, four counties were hand-picked for oversampling: Nairobi (the capital of Kenya), Nyandarua (urban area, with relatively poor educational performance despite its low poverty rate), and Nyamira and Siaya (both as interesting

comparison cases).[6] For the 2018 SDI health surveys in Sierra Leone, all hospitals and health centers were selected within a stratum, whereas other types of facilities like clinics and health posts were randomly sampled.

Overall, sampling procedures are carefully planned, with protocol specified ahead of data collection activities. Contextual decisions to select nonrandomly or eliminate certain geographic locations or types of facilities are taken ex ante and in a fashion that does not compromise rigor and representativeness of results or comparability within and across countries. This process ensures that all indicators have a known and acceptable margin of error and can be used for monitoring and evaluation purposes.

Notes

1 | Sample selection is said to be stratified if some form of random sampling is applied separately in each set of distinct groups formed from all of the entries on the sampling frame from which the sample is to be drawn (Lavrakas 2008). During the stratification process, all sampling units are recategorized in mutually exclusive groups ("strata") using a combination of features such as region (for example, province), area (for example, urban-rural), type of provider (for example, public, nonprofit, private), or others.

2 | Lists of all school and health center facilities (and information on the characteristics of these facilities) are usually obtained from the ministries (education and health) or the national statistics office. After obtaining these lists, a final sample frame is generated using additional sources in a data-cleaning process to avoid missing entries, duplicates, and outdated or incorrect information.

3 | In some cases, this second stage becomes the first stage of sampling when schools or facilities are drawn directly from each defined stratum, skipping the selection of cluster locations as a first step (for example, the 2015 SDI education survey in Niger).

4 | Steps to generate backup units for surveyed individuals (staff and pupils) are also planned during sampling. For instance, for the 2015 SDI education survey in Niger, a risk of pupils not returning after lunch was identified during pretesting. Consequently, field teams were instructed to learn from the teacher which sampled pupils might not return after lunch in order to minimize the risk of truncating the lower tail of the performance distribution. Additionally, teams also drew three extra pupils who were kept in reserve in case one of the 10 originally sampled pupils was not available.

5 | Health facility weights were estimated for the following SDI surveys: Kenya (2012, 2018), Madagascar (2016), Niger (2015), Nigeria (2013), Senegal (2010), Tanzania (2014, 2016), Togo (2013), and Uganda (2013). Health care provider weights were calculated for the following SDI surveys: Kenya (2018), Madagascar (2016), Niger (2015), Nigeria (2013), Tanzania (2014, 2016), Togo (2013), and Uganda (2013). School weights were calculated and included in the data sets of the following SDI surveys: Kenya (2012), Madagascar (2016), Mauritania (2016), Mozambique (2014), Niger (2015), Nigeria (2013), Senegal (2010), Tanzania (2010, 2014, 2016), Togo (2013), and Uganda (2013). Finally, pupil and teacher weights were calculated for the following SDI surveys: Madagascar (2016), Mauritania (2016),

Mozambique (2014), Niger (2015), Nigeria (2013), Tanzania (2014, 2016), Togo (2013), and Uganda (2013).

6 | Both counties are located in Nyanza Province, are relatively rural, and have high poverty rates. However, both also have very different educational performance in terms of passing grades and thus presented good cases for a comparison analysis.

Reference

Lavrakas, P. J. 2008. *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage Publications. https://doi.org/10.4135/9781412963947.

Appendix C: Survey methodology

The Service Delivery Indicators (SDI) initiative aims to provide national and subnational evidence on the quality of primary education and basic health service delivery from the perspective of an average citizen accessing these services. To do so, SDI survey instruments are structured meticulously to collect information on three aspects of service delivery: (1) provider effort, (2) provider knowledge, and (3) facility inputs. Provider effort is measured by collecting data on teachers' and health workers' absenteeism rates, health care providers' caseloads, and teachers' time spent in teaching activities. Provider knowledge is measured by collecting data on teachers' knowledge of the curriculum and quality of their pedagogy as well as health workers' diagnostic and treatment accuracy, adherence to clinical guidelines, and management of maternal and neonatal complications. Facility inputs are measured by observing each facility's equipment and infrastructure availability, each health facility's medicine availability, and each school's student-to-textbook and student-to-teacher ratios.

Survey teams visit schools and health facilities to collect data through a combination of observation and interview techniques.[1] SDI survey instruments are nimble and use cutting-edge data collection methods, allowing for relatively

rapid fieldwork and data analysis. In addition, SDI projects involve rigorous quality control with defined standards. Through a participatory approach during planning and design, World Bank staff, working alongside local technical teams and other stakeholders, ensure that survey modules reflect national standards while maintaining cross-country comparability. During the implementation of data collection activities, enumerators receive comprehensive training and follow established protocols, and trained supervisors randomly check fieldwork. SDI analysis follows internationally accepted definitions of indicators and uses reproducible statistical programming.

The overall duration of an SDI project depends on the sample size and country context. As a rough estimate, an SDI survey can take between 12 and 18 months, depending on procurement, planning, and seasonality. Its project cycle encompasses four stages: (1) survey planning and design, (2) survey implementation, (3) analysis and capacity building, and (4) dissemination and mobilization.

During the planning and design stage, the World Bank's SDI core team (hosted at the Human Development Practice Group's Chief Economist Office), task teams, and government authorities engage in initial conversations to understand the SDI survey's goals and the processes involved, identify the country's priorities and capacity needs, and align the SDI survey with the country's sector strategy. These initial conversations involve virtual and in-country consultations and, potentially, field visits to facilities.

After the SDI survey's overall goals are defined and agreed upon, the SDI core team, task teams, and a local technical working group collaborate to contextualize and adapt the survey instruments[2] and define sampling procedures. The local technical group includes ministry and regional officials, development partners, and colleagues from research and academic institutions. Survey protocols and instruments are submitted to the corresponding authorized ethics committee or institutional review board for approval. Once survey instruments are finalized and approved, a computer-assisted personal interviewing designer is hired to program the instruments onto tablets. Additionally, a field coordinator is assigned to supervise and coordinate the overall data collection fieldwork, and a survey firm is hired to conduct data collection activities.

During this planning and design stage, roles and responsibilities among the different stakeholders involved in the SDI project are also defined and agreed upon. In general, the SDI core team's main responsibility is to ensure that a high-quality survey can take place, guaranteeing indicators that are nationally representative and comparable across time and countries. The core team's assigned functions often include providing (1) standardized materials and suggestions for survey adaptation, (2) technical guidance on survey design and sampling procedures, (3) protocol for and assistance in quality control, and (4) guidance on training, planning, and procurement (when needed).

Task teams are responsible for the overall coordination of the project, ensuring that operations run smoothly during the different stages of the work. Task teams lead the dialogue with government authorities and identify the local technical working group. In addition, task teams are responsible for contracting and managing the survey firm, taking the lead in analysis and report writing, and creating a dissemination strategy for the SDI results.

Governments provide a unique perspective about the country's contextual challenges, opportunities, and needs. They also promote an appropriate environment that ensures compliance with SDI survey quality standards and buy-in from national and subnational stakeholders.

Finally, the fieldwork team (field coordinator, survey firm, and information technology [IT] technicians) agree on a contract to comply fully with survey protocols and guidelines when conducting data collection activities. During the survey implementation stage, the survey firm's supervisors and enumerators are trained for approximately three weeks with a combination of workshop and field-based sessions. In parallel, instruments are piloted and finalized for approximately two weeks. Once data collection activities begin, they take approximately eight weeks to be completed, depending on the sample size and logistical considerations.

The survey firm conducts these data collection activities by deploying several survey teams to the field. Survey teams are usually composed of a field team supervisor, an IT technician, and two to three enumerators. The field coordinator tracks and oversees work conducted by the survey firm and its survey teams and communicates overall progress to task teams daily. The field coordinator is responsible for overall data quality control during the implementation stage.

The field team supervisor is the senior member of each field team. He or she is responsible for the well-being and safety of team members, completion of the assigned workload, and maintenance of data quality for that team. Each field team supervisor receives his or her assignments from and reports to the field coordinator.

The responsibilities of the enumerators include completing all required interviews, observations, and assessments or tests. Enumerators are also expected to check completed survey forms to ensure that all questions were asked and responses were properly recorded. This check should be done at the end of the first visit to allow for later completion of any questions that were inadvertently skipped or for which responses were incorrectly recorded. Enumerators upload their data as frequently as possible to allow for real-time checking of results. Finally, IT technicians are employed to provide technical support to the survey teams, helping with troubleshooting tablet and software malfunctions during the fieldwork.[3]

Before the fieldwork starts, introduction letters are sent to all facilities in the sample to ensure the timely flow of information and preparations for smooth and efficient data collection. Apart from requesting permission and collaboration from the facility to allow the field team to conduct the survey, these letters also detail the scope of the survey and preparations that facilities must have in place before the arrival of the field team. The letters also specify the number of visits that field teams will make to the sampled facilities.[4] Once survey teams arrive at the facilities, field team supervisors, enumerators, and IT technicians carry special identification documents. They bring copies of introduction letters and official letters sent to facilities by their corresponding governing body (for example, ministry) to introduce the survey.

Data collection activities consist of two visits to each sampled facility: an announced visit following by an unannounced visit (an example on how survey visits are usually structured is shown in table C.1).

During the announced visit, enumerators collect data on the following:

- The facility's inputs and procedures
- The full roster of teachers or health workers and their qualifications
- Professional knowledge of up to 10 randomly selected teachers (measured using fourth-grade tests and pedagogical assessments) or health workers (measured using patient case simulations on symptoms of high-burden diseases)
- Knowledge of up to 10 randomly selected pupils in math, language, and nonverbal reasoning (in schools)
- Classroom observation for 45–60 minutes of one randomly selected lesson (only for SDI education surveys)
- Perceptions and satisfaction of visiting clients (only in health facilities and for some countries)

During the unannounced visit, which occurs at least five days after the announced visit, enumerators collect data on the following:

- Attendance check of presampled teachers or health facility workers
- Unstaffed classrooms (only for schools)[5]

Detailed coordination and specialized skills are required to implement SDI surveys successfully. Therefore, survey firms comply with strict selection and training processes when they hire enumerators and assemble survey teams. The "patient case simulations" (or "clinical vignettes") are a good example of the high level of coordination and specialization required by teams to implement SDI surveys. Patient case simulations are answered by health care providers (who provide care for patients) to measure their knowledge and competence in diagnosing and treating key illnesses and complications. To conduct this section of the survey successfully, at least two enumerators (often with previous medical knowledge and expertise) are assigned to a survey team and receive intense training on a detailed set of instructions that must be followed during fieldwork

TABLE C.1 Example of a typical SDI education survey instrument

Visit and instrument model	Module title	Data collector	Interviewee	Description
First visit				
Module 1	School information	Enumerator 1	Principal or head teacher	Administered to the head of the school to collect information about the type of school, school facilities, school governance, number of students, and school hours
Modules 2A and 2B	Teacher roster	Enumerator 1	2A: Head teacher 2B: Selected teachers	2A: Administered to head teacher to obtain a list of all school teachers 2B: Administered to selected teachers to collect information about teacher characteristics
Module 3	School finances	Enumerator 1	Principal or head teacher	Administered to the head of the school to collect information about school finances
Module 4	Classroom observation	Enumerators 1 and 2	Observation (teachers and pupils)	An observation module to assess teaching activities and classroom conditions
Module 5	Pupil test	Enumerator 2	Pupils	A test of students to have a measure of student learning outcomes in mathematics and language in fourth grade
Module 6	Teacher assessment	Enumerator 2	Teachers	A test of teachers covering mathematics and English subject knowledge and teaching skills
Second visit				
Module 2B	Teacher roster	Enumerators 1 and 2	2B: Selected teachers	Administered to selected teachers to measure absence rates and count number of unstaffed classrooms

Source: Service Delivery Indicators (SDI) education survey for Kenya (2012).

activities. Once at health facilities, each enumerator is assigned a specific role to perform during the simulation activity. One enumerator, the "patient enumerator," acts as the patient (or the caregiver of the patient) and presents the symptoms of each of the clinical simulation cases to the health worker. A second enumerator, the "observer enumerator," instructs the health worker to manage the clinical simulation case as he or she would a real patient, assuming that the health facility is fully functional and equipped, staffed, and stocked with all medications and supplies as per national guidelines. The observer enumerator silently notes the clinical questions and procedures provided by the health worker during the simulation exercise, while the patient enumerator provides the associated standardized response for each of the health worker's clinical questions and procedures.

Distinct task assignments are also distributed to enumerators during SDI school visits. For instance, during the first, announced visit, one enumerator

could interview the principal or head teacher to complete the school infor-mation module (module 1), collect the list of all teachers (module 2A), and complete the school finances module (module 3); another enumerator could observe class lessons (module 4) and administer tests to pupils and teachers (modules 5 and 6, respectively).

Before leaving the facility, the field team supervisor reviews each completed interview and observation module to check for completeness and consistency. If the instruments are incomplete, the enumerator is sent back to collect the missing data from the respondent. Special efforts are made to avoid high rates of nonresponse. If the respondent refuses to be interviewed, a reasonable and polite effort is made to elicit consent from the respondent to participate in the study. Because refusals may stem from misconceptions about the survey or other prejudices, enumerators are carefully trained and encouraged to consider the respondent's point of view, adapt to it, and reassure him or her.

The field coordinator then implements a second level of quality control, often using spot-check and back-check techniques to ensure the quality and legitimacy of data collected by survey teams. During these fieldwork checks, the field coordinator accompanies a survey team and provides feedback and recom-mendations to enumerators to improve the delivery of the instrument. Finally, task teams and local technical teams provide the final stage of quality control through periodic review of preliminary data, conducting high-frequency anal-ysis to ensure smooth conduct of survey activities and recommend timely rec-tifications, if needed.

During the third stage, analysis and capacity building, two-week workshops are conducted to build the technical working group's capacity for analyzing and interpreting data and undertaking preliminary analyses of core indicators. Tables of key results and slides of summary results are drafted for preliminary review by the task team and government authorities, and a draft report is written.

Finally, during the dissemination and mobilization stage of the SDI project, data and findings are made broadly available to inform the development of pol-icies and design of interventions to improve health and education outcomes in the client country and around the globe. High public awareness of SDI results mobilizes policy makers and citizens as well as donors and other stakeholders into action. SDI projects use a multitude of dissemination materials targeting a variety of consumers of information, including a standardized SDI report, slide decks with key results, and an SDI results brief.

To date, more than 3,200 schools, 34,000 teachers, 7,500 health facilities, and 66,000 health workers have been surveyed in 24 SDI surveys completed across Africa. Now the SDI initiative is expanding globally, and, through fiscal 2021/22, SDI survey work will advance in several countries in Africa and other regions. Participating countries are at different stages of the SDI process, rang-ing from procurement to data collection and analysis.

Notes

1 | SDI facility surveys are increasingly coupled with household surveys to provide a complete perspective on the quality of basic health and primary school services.

2 | Technical adaptations of the instruments are carried out—for instance, aligning the list of essential medicines and medical equipment to the national guidelines.

3 | In countries where more than one language is spoken, local translators are hired to accompany survey teams.

4 | For the second visit, which is expected to be a surprise visit, the letter should not specify when the visit will occur.

5 | Often, enumerators take advantage of this second visit to complete any information missing from the first visit.

Appendix D: Methodological groundwork for the SDI teacher and student assessments

The Service Delivery Indicators (SDI) teacher and student assessments are based on a study commissioned by the SDI program and completed in 2010 by Andrew Cunningham, David Johnson, and Rachel Dowling (Cunningham, Johnson, and Dowling 2012). Their study reviewed, analyzed, and summarized the national third- and fourth-grade curricula of 12 countries in order to identify common themes (for example, problem solving, long division): Botswana, Ethiopia, The Gambia, Kenya, Madagascar, Namibia, Nigeria, Rwanda, the Seychelles, South Africa, Tanzania, and Uganda. The study focused on Africa,

because this is the region in which the SDI program was conceived. Study results informed the content and question format used for the teacher and student assessments in SDI education surveys.

The SDI teacher assessment was designed with the objective of examining whether teachers have the basic reading, writing, and arithmetic skills that lower-primary students need in order to progress in their education. A score of 80 percent correct is interpreted as the minimum knowledge required for a teacher to be effective, which becomes an SDI indicator representative at the national level. The tests also examine the extent to which teachers demonstrate mastery of skills that are above the level they are teaching (but still at the level of primary school content).

The SDI teacher assessment includes two sections, each lasting 35 minutes. The teacher knowledge section resembles grading a math and literacy exam. It asks teachers to correct a letter with grammatical errors, fill in missing words in sentences, solve arithmetic problems, read the time on a clock, and read and understand simple graphs. All of the content is based on material drawn from the national curricula of the 12 reference countries. Similarly, the pedagogical section resembles tasks that primary teachers face daily: preparing to teach a lesson, assessing differences in children's abilities as learners, and correctly evaluating students' learning achievements and progress. Because of the nature of the content, teacher assessments are paper based and implemented in groups at an arranged time when teachers do not have to worry about other responsibilities (for example, during lunch break or after classes). Unfortunately, only teachers who are present in the school can be assessed, so the results represent an upper boundary if it is assumed that teacher absence and knowledge are negatively correlated.

The SDI student assessment includes three sections: mathematics, language, and nonverbal reasoning. For the language section, the language of instruction in a given country is always tested. However, in some countries, the vernacular or other common languages are also tested in a subsample of students. Exams are always translated by the local implementation team and validated by government counterparts. The language questions include items on correctly identifying a letter or word, matching a word with an image, reading a sentence and a paragraph out loud, and answering questions based on a simple story. The nonverbal reasoning section includes four questions on completing patterns following Raven's Progressive Matrices. The mathematics section is the most extensive and covers items on identifying and ordering numbers, adding, subtracting, and multiplying up to three digits, dividing up to two digits, handling fractions, completing a word problem needing arithmetic, and completing a sequence of numbers. Because of the multiple languages tested and the small number of nonverbal items, the math section is the most comparable across countries.

Unlike the teacher assessment, the SDI student assessment was initially designed only to serve as a correlate for other SDI indicators, not to result in an SDI indicator nor to serve as a large-scale international assessment. The content of the student assessment was also derived from Cunningham, Johnson, and Dowling (2012) and includes topics covered in the third- and fourth-grade national curricula of the 12 countries. Each student has 25 minutes to complete the test, including 10 minutes for the language section, 10 minutes for the mathematics section, and 5 minutes for the nonverbal reasoning section. If the student has not completed a section after the allocated time, the test proctor makes the student move on to the next section. The student assessment is implemented orally and is conducted one-on-one to prevent guessing and missing information from students who are unable to read. For the written component, the proctor informs the student that he or she can answer questions in any order.

In contrast with the teacher assessment, the time of the year in which the SDI student assessment is implemented might affect the results, particularly because the items are drawn from the third- and fourth-grade curricula and the assessment is taken by students somewhere in the middle of the fourth-grade school year. Although this is important for interpreting observed differences, SDI students' assessments are not the only international assessments that suffer from this caveat, and informative cross-country comparisons are still useful.

Given its solid grounding in the national curricula of multiple countries, the breadth of topics covered, and the careful implementation guidelines followed, the SDI student assessment results prove to be robust to several psychometric properties. SDI tends to be implemented in countries where information about the education sector is scant and no other international assessments have been implemented. The SDI team is currently working with the Harmonized Learning Outcomes team to have SDI assessments results included in this database, in instances where the results are useful and no better information is available. This effort will help to improve the precision and increase the availability of other World Bank Group initiatives such as the Human Capital Index.

Reference

Cunningham, A., D. Johnson, and R. Dowling. 2012. "Teaching Standards and Curriculum Review." Report 1-31072012, World Bank Group, Washington, DC.

Have teachers mastered the subject matter they are teaching? Can doctors accurately diagnose and treat critical health conditions? Are schools and health facilities sufficiently stocked with needed equipment and supplies? Are they sufficiently supported and staffed to optimize learning and health care outcomes?

For the past decade, the World Bank's Service Delivery Indicators (SDI) surveys have collected nationally representative data in countries across Africa to answer these questions. The surveys aim to measure the quality of services where they meet citizens: in schools and health facilities.

The Quality of Health and Education Systems Across Africa: Evidence from a Decade of Service Delivery Indicators Surveys identifies areas of achievement and constraint in service delivery, shedding light on how service delivery may foster or stunt human capital accumulation. SDI surveys show that schools and health clinics across Africa are still falling short in some critical areas.

The delivery of primary care services is very heterogenous between and within countries. Many health facilities lack the basic necessities to provide proper care, such as essential medicines, basic diagnostic equipment, and adequate water and sanitation. Moreover, health care providers' ability to diagnose and treat common health conditions correctly is low and distributed unevenly. Health personnel's absence from health facilities remains a concern across the surveyed countries.

Learning is low, and, not unlike health care, levels of student learning vary significantly across countries: less than half of grade 4 students can recite a simple sentence or perform basic mathematical operations. This deficient learning is correlated with teachers' low levels of content knowledge and sub-par pedagogy skills. Some schools are also missing crucial inputs, such as blackboards or private and gendered toilets, and struggle with high pupil-teacher ratios.

Despite these challenges, success stories in both sectors illustrate the quality of service delivery that could be achieved and showcase the dedication of teachers and medical staff across Africa. By studying data from thousands of facilities, considering the local context, and drawing insights from the literature, this book offers important insights for how countries can strengthen health and education systems and build back better in the wake of the massive disruptions brought about by the COVID-19 pandemic.

WORLD BANK GROUP