WORLD BANK GROUP

THE QUALITY OF HEALTH AND EDUCATION SYSTEMS ACROSS AFRICA

Evidence from a Decade of Service Delivery Indicators (SDI) Surveys

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

1818 H Street NW, Washington DC 20433

Telephone: 202-473-1000; Internet: www.worldbank.org

THE QUALITY OF HEALTH AND EDUCATION SYSTEMS ACROSS AFRICA

Evidence from a Decade of Service Delivery Indicators (SDI) Surveys

Roberta Gatti

Kathryn Andrews

Ciro Avitabile

Ruben Conner

Jigyasa Sharma

Andres Yi Chang

WORLD BANK GROUP

Photo Credits

Page 5	Sarah Farhat, World Bank
Page 11	Dominic Chavez, World Bank
Page 13	Henitsoa Rafalia, World Bank
Page 21	Dominic Chavez, World Bank
Page 47	John Hogg, World Bank
Page 56	Dominic Chavez, World Bank
Page 73	Dominic Chavez, World Bank
Page 80	Dominic Chavez, World Bank
Page 91	Sambrian Mbaabu, World Bank
Page 95	Sarah Farhat, World Bank

TABLE OF CONTENTS


cook - cooked
wash - washed
dance - danced
Jumped
watched
The day befo
Last week
Last year
Last month

ACKNOWLEDGEMENTS

This booklet stems from the collaborative effort of a core team led by Roberta Gatti and including Kathryn Andrews, Ciro Avitabile, Ruben Conner, Jigyasa Sharma, and Andres Yi Chang. Chapter 2 of the booklet draws significantly on Andrews et al. (2021). The team is grateful to Michael Crawford for the box on language of instruction; to Melissa Adelman for describing the SDI experience in the Democratic Republic of Congo; to Maud Juquois, Rija Andriantavison, Anna Olefir, and Adria Rakotoarivony for the description of the SDI experience in health/education in Madagascar; to Pamela Mulet for the impacts of SDI data in Niger; to Nicola Dehnen for her contributions to the early stages of the education analysis; and to Sebastian Insfran and Juan Mejalenko for analytical support. Sebastian Insfran also co-authored the technical appendices. Alexander Irwin provided excellent editorial support. Nicola Dehnen and Shuqiao Sun read the draft booklet carefully and provided detailed feedback.

We are indebted to Christophe Rockmore and Waly Wane for their leadership and ongoing support of the Service Delivery Indicators initiative. With their seminal work on *World Development Report 2004: Making Services Work for the Poor,* Ritva Reinikka (Aalto University) and Shanta Devarajan (Georgetown University) laid the groundwork for the innovative way in which the Service Delivery Indicators have measured the quality of health and education at the point where the service meets the citizen. Their continuous guidance has been invaluable. Jishnu Das (Georgetown University) played a key role as an external advisor throughout the booklet's drafting and editing process. Deon Filmer has been generous with technical and strategic advice towards the development of the Service Delivery Indicator initiative as a whole.

This booklet benefited from consultations with the Africa Human Development leadership team and from consultations held at the World Bank Nairobi Country Office (October 2019). Victor Alegana (KEMRI-Wellcome) and researchers from KEMRI-Wellcome Trust Research Programme and Strathmore University, Nairobi, provided useful comments. We are grateful to our peer reviewers Melissa Adelman, Patrick Eozenou, Deon Filmer, Elisabeth Huybens, and Jeremy Veillard for their thoughtful guidance and also appreciate the comments of Clara Delavallade, Caren Grown, and Eliana Rubiano on behalf of the Gender Group, as well as of Hana Brixi, Ramesh Govindaraj, Igor Kheyfets, and Laura Rawlings on behalf of the Human Capital Project team. Additional thanks go to Omar Arias, Joao Pedro Azevedo, Marta Carnelli, Damien de Walque, Koen Geven, Sergio Venegas Marin, Ezequiel Molina, Diego Armando Luna Bazaldua, Christophe Rockmore, Halsey Rogers, and Noah Yarrow for detailed comments on an earlier draft of the booklet. At different stages of this work, the team benefitted from feedback from Jaime Saavedra, Jed Friedman, Magnus Lindelow, and Feng Zhao.

This booklet was completed under the strategic leadership of Mamta Murthi.

LIST OF ABBREVIATIONS

ACT	Artemisinin combination therapy
AERC	African Economic Research Consortium
CAPI	Computer-Assisted Personal Interviewing
CHW	Community health worker
CO	Clinical officer
CRP	c-reactive protein
D-WMS	Development World Management Survey
DALY	Disability-adjust life year
DHIS2	District Health Information Software 2 (Kenya)
DHS	Demographic and Health Surveys
DRC	Democratic Republic of Congo
EMIS	Education management information system
EQUIP	Education Quality Improvement Project
ESP	Education sector plan
ESR	Erythrocyte sedimentation rate
GDP	Gross domestic product
GFF	Global Financing Facility for Women, Children and Adolescents
GLAD	Global Learning Assessment Database
HbA1c	Hemoglobin A1c test for diabetes
HCI	Human Capital Index
HLO	Harmonized Learning Outcomes
HMIS	Health management information system
IBRD	International Bank for Reconstruction and Development
ICT	Information and communication technology
IEA	International Association for the Evaluation of Educational Achievement
IHME	Institute for Health Metrics and Evaluation
ILSA	International large-scale assessment
IMCI	Integrated Management of Childhood Illness
IPC	Infection prevention and control
IRT	Item Response Theory
IT	Information technology
IV	Intravenous
L1	An individual's first language learned ("mother tongue")
LIRE	Learning Improvement for Results in Education (Niger)
MELQO	Measuring Early Learning and Quality Outcomes
mhGAP	Mental Health Gap Action Programme (World Health Organization)
MLE	Maximum likelihood estimation
NCD	Non-communicable disease
NG	Nasogastric
NGO	Non-governmental organization
NLSA	National large-scale assessment
OECD	Organisation for Economic Co-operation and Development
ORS	Oral rehydration salts
PCA	Principal component analysis
PHC	Primary health care
PHCPI	Primary Health Care Performance Initiative
PIRLS	Progress in International Reading Literacy Study
PISA	Program for International Student Assessment
PPE	Personal protective equipment
PPP	Purchasing power parity
PPS	Probability proportional to size
PSU	Primary sampling unit
RDT	Rapid diagnostic test
SABER	Systems Approach for Better Education Results
SD	Standard deviation
SDI	Service Delivery Indicators
SEHC	Satisfaction of Employees in Health Care
TB	Tuberculosis
UHC	Universal health coverage
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development
WaSH	Water, sanitation, and hygiene
WBG	World Bank Group
WDR	World Development Report
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 high-quality services. Yet even before the COVID-19 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 Indicator (SDI) surveys supply this evidence by measuring key aspects of the quality of education and health services. SDI surveys collect data at the facility level – in schools, clinics, and hospitals – sampling anywhere between 250 and 3,000 public and private facilities in urban and rural areas of each country. 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.

SDI surveys focus on the experience of average citizens in primary health care and primary schools. The surveys collect data on four 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 student learning outcomes by testing fourth-graders' language, math, and non-verbal skills.

A DECADE OF DATA SHOWS 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 booklet also charts directions for a measurement agenda to inform how countries develop policies, design interventions, and deliver essential services.

LEVERS FOR HEALTH SYSTEM CHANGE: STRENGTHEN SKILLS AND TOOLS ON THE FRONT LINES

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 highlight patterns in service delivery, identifying both the strengths and 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 provider exceeding 20 percent in almost all countries, although the majority of this absence is authorized. These data also put numbers on a much-debated issue of efficient human resource allocation. Across all facilities, the average health care provider in these countries sees 13 patients per day, varying from a low of three 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 optimize service delivery. Absences, which are higher at public facilities, could potentially be reduced with better supervision and management. To help overburdened facilities, 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 persist. Finally, infrastructure gaps are particularly pronounced at rural facilities, and governments can continue to close this gap 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 recite 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, a large share are not present in the classroom during their assigned teaching times. Teacher's skills are also low, both in 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 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 national policy dialogue from inputs towards 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 2014 SDI Education Report. 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 in 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 children's nutrition and growth trajectories and common non-communicable 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 policymakers.

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 the continuous innovation in the SDI measurement platform offer important contributions toward achieving this goal.

CHAPTER 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.[1]

Even before the COVID-19 pandemic, many countries were struggling with low human capital outcomes. The World Bank's 2020 Human Capital Index (HCI) 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.[2] This statistic reflects decades of progress in extending life expectancy and expanding schooling access and quality, but 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.[3] Failure to make these timely investments will make it harder and costlier to reduce inequalities in the future.

COVID-19 IS 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 in human lives, but also through potential long-term reductions in healthy functioning among survivors of the disease.[4] 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.[5] Along with these direct impacts, the pandemic is also affecting systems and people's lives through indirect mechanisms. These include disruptions in non-COVID-19 essential health services and sharp declines in family income following lockdowns, among other pathways.[6]

1 World Bank (2020d) and Flabbi and Gatti (2018).
2 World Bank (2020d).
3 Kim (2018).
4 Mahase (2020).
5 World Bank (2020c).
6 Corral and Gatti (2020).

The pandemic has impacted both the supply and demand for essential care, as providers have struggled to keep up with the burden of care for COVID-19 patients; 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.[7] Evidence already indicates sharp drops in routine immunization coverage during COVID-19.[8] 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 and 45 percent, respectively, attributable to COVID-19-related poor nutrition and interruption of essential health services.[9]

On the education side, there has been concern for a potential "lost generation" of learners, reflecting the direct threat of the crisis, the prolonged interruption of schooling, and expected increases in poverty.[10] 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.[11]

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 to 40 million additional people pushed into extreme poverty in 2020, disruptions in immunizations and other routine health services,[12] and widespread school closures,[13] the COVID-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 it even more imperative that resources are used wisely.

7 Elston et al. (2017).
8 Jain et al. (2020).
9 Roberton et al. (2020).
10 UNICEF (2020).
11 World Bank (2020c).
12 WHO (2020a).
13 Human Rights Watch (2020).

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 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."[14] 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 critical care capacity and guidance on isolation/containment procedures. In addition, many lower-level facilities lacked adequate handwashing facilities, had little or no personal protective equipment (PPE) and did not have necessary communication equipment.[15] Similarly, education systems have been strained by the abrupt shift to remote learning and have had to quickly make 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 prepared to develop human capital and meet citizens' expectations in the long term. As COVID-19 vaccines are being rolled out, governments can look towards a post-pandemic future and incorporate lessons from the crisis to build greater resilience into 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 highest returns. This also means that measurement will play a crucial role in enabling countries to transform their health and education systems post-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.[16] Reciprocally, 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.

14 USAID (2012).

15 Sharma et al. (2020).

16 Currie (2009).

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.[17] 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**). By documenting providers' competence and behavior, the availability of inputs, and – in the case of education – children's learning, SDI

BOX 1: WHAT DO SDI SURVEYS MEASURE?

WHAT DO PROVIDERS KNOW? SDI surveys measure teachers' knowledge directly by assessing how teachers answer questions on the 4th-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. The innovative methodology used for the vignettes makes the SDI surveys the only surveys of this scale that measure health care provider knowledge in a comparable way across countries.

WHAT DO PROVIDERS DO? Visiting unannounced, SDI survey enumerators assess the levels of absence among teachers, observing if there are classes unattended; and among health workers, cross-checking the presence of providers vis-à-vis 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 not only limited by the technical capacity and behavior of providers. It is also constrained 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 properly store those vaccines. 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 different doctors, in a different clinic. Hence SDI surveys do not include measures of health outcomes. In contrast, the link between quality and knowledge among teachers 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.

17 The Systems Approach for Better Education Results (SABER) is a WBG 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 systematically strengthen their education systems and the ultimate goal of promoting Learning for All.

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 engaging a broad set of stakeholders – from governments to trade unions and parents' and patients' groups.

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 (AERC) and later supported by the William and Flora Hewlett Foundation and the African Development Bank, worked together to develop novel survey tools and methodology to comprehensively measure 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, anti-malarial 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 Grade 4 children are 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 multi-faceted 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*.[18] This framing emphasizes the idea that measurement and information on providers' performance enables 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 by an active process of

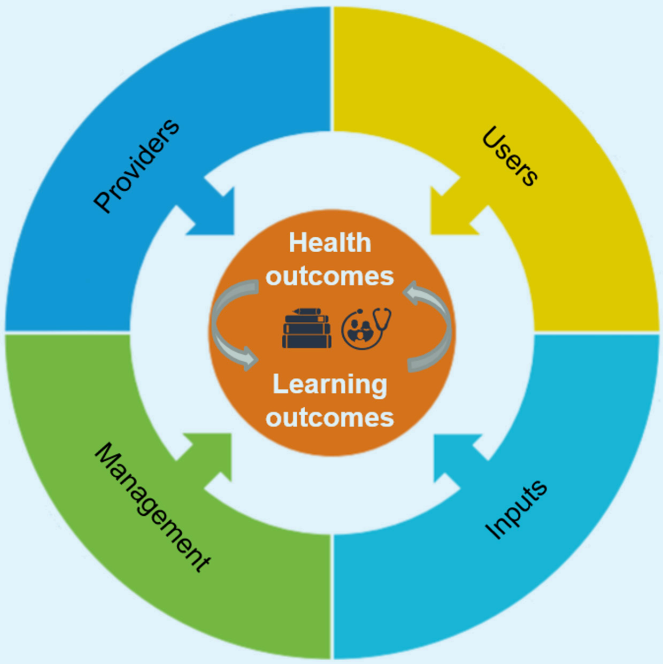
18 World Bank (2004).

citizen engagement. This principle is echoed in the 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."[19]

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 effectively and fairly distributed. 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 bring about improvements.

The relevance of the information provided by SDI surveys depends on how well what is 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.[20] On the health side, the Lancet Global Health Commission on High Quality Health Systems developed a framework showing how processes, impacts, and inputs relate.[21] The Lancet Commission schema includes more detail than the *World Development Report* (WDR) 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**. While 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

FIGURE 1 INPUTS AND OUTCOMES IN HEALTH AND EDUCATION



Source: Adapted from World Bank (2018).

19 Kruk et al. (2018a).

20 World Bank (2018).

21 The Lancet Global Health Commission on High Quality Health Systems (2018) proposed 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 WDR elements of providers, inputs, management, and users. The Lancet framework also includes three components under quality impacts: better health, confidence in the system, and economic benefit, which are summarized as outcomes in the WDR framework.

sector, the legal system, and civil society organizations, among others. Nonetheless, these four main categories of the concise version align well 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 around measuring the quality of management and other key drivers of service quality (see Chapter 4).

AIMS AND STRUCTURE OF THE BOOKLET

This booklet documents lessons learned from a decade of SDI surveys.[22] 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 booklet 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 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 (PHC). This chapter 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 health care provision.

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. While 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, while other changes are premised on better understanding the context in which health and education services are utilized. 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.

The booklet concludes with Chapter 5, which indicates 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 booklet is intended to help 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 this booklet aims to provide a measurement-driven contribution to the policy discussions that are now being engaged.

Accompanying the analytical work of this booklet, 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 interactively visualize the SDI health and education indicators for all countries included in the booklet. The website, data repository, and data visualization tool will be updated periodically as more countries complete SDI surveys.

22 This booklet 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 Simeon Djankov and Ugo 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: 60–70.

Flabbi, L. and R. Gatti, (2018). "A Primer on Human Capital." Policy Research Working Paper No. 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, 31 August 2020.

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/

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

Kim, J. 2018. "The Human Capital Gap." *Foreign Affairs* July/August 2018.

Kruk, M. E., A.D. Gage, C. Arsenault, K. Jordan, H.H. Leslie, S. Roder-DeWan, O. Adeyi, P. Barker, B. Daelmans, S.V. Doubova, M. English, E.G. Elorrio, F. Guanais, O. Gureje, L.R. Hirschhorn, L. Jiang, E. Kelley, E.T. Lemango, J. Liljestrand, A. Malata, … and M. Pate. 2018a. "High-quality health systems in the sustainable development goals era: time for a revolution." *The Lancet Global Health.* 2018;6(11): e1196–ee252. doi: 10.1016/S2214-109X(18)30386-3.

Mahase, E. 2020. "Covid-19: What do we know about 'long covid'?" *British Medical Journal* 2020;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." *The Lancet Global Health* 2020 Jul;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 No. 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. 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. United States Agency for International Development, Washington, DC.

WHO (World Health Organization) 2020a. "Pulse survey on continuity of essential health services during the COVID-19 pandemic: interim report, 27 August 2020." World Health Organization, 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. 2020c. "The COVID-19 Pandemic: Shocks to Education and Policy Responses." Washington, DC: World Bank.

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

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

CHAPTER 2

HEALTH SERVICE DELIVERY IN NINE AFRICAN COUNTRIES

BACKGROUND: AN OPPORTUNITY TO TRANSFORM PRIMARY HEALTH CARE

One of the most basic human aspirations is the hope for a long and healthy life. When health problems do arise, people want to access 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).[23] For four decades, PHC reform has been at the center of global efforts to expand population access to health services and to improve health outcomes. PHC values have been reflected in the Millennium Development Goals and Sustainable Development Goals.[24] Recently, country health leaders have reaffirmed their recognition that achieving universal health coverage (UHC) depends on governments' ability to expand PHC access to marginalized groups while ensuring high-quality care for all.[25]

COVID-19 has tested health systems in numerous ways. The pandemic has also confirmed what many in the health community already knew – that strong primary health care systems are the foundation of population health during and beyond health crises.[26] PHC systems are vital to control local outbreaks through prevention of transmission, detection of cases, and quick clinical response. In addition, PHC systems need to ensure the continued delivery of non-pandemic essential health services to communities.[27]

COVID-19 has also revealed weaknesses in many countries' PHC systems and aggravated existing inequities. For instance, preliminary evidence from 29 countries suggests decreased access to sexual and reproductive health services and increased violations of related rights, as an indirect impact of COVID-19.[28] As in previous epidemics, there is a risk of substantial mortality attributable to the disruption to routine PHC services.[29] Modeled scenarios suggest that disruption of PHC services may cause more deaths in some settings than COVID-19 itself.[30] 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, there will be a unique

23 UNICEF and WHO (1978).
24 United Nations (2011); UN General Assembly (2015).
25 World Health Organization (2013).; WHO, OECD and World Bank (2018).
26 WHO (2020).
27 PHCPI and World Bank (2020).
28 Endler et al. (2020).
29 Parpia et al. (2016).
30 Roberton et al. (2020).

opportunity to strengthen and reform PHC, increasing investment while making it more fit for purpose. These reforms can be far reaching and will offer a chance to reimagine the health sector, strengthen linkages within it, and address neglected challenges. Measuring and ensuring quality remains vital for health-systems improvement, as estimates indicate that there are more deaths globally due to low quality of health care (5 million deaths annually) than to lack of access to care (3.6 million deaths annually).[31] To achieve the greatest population health gains, 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 insight to advance health system reforms, with actionable indicators and a roadmap for improvement.

SDI HEALTH SURVEYS: A FINGER ON THE PULSE OF PHC

The SDI health facility survey offers 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 effort towards 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 different 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 health system performance indicators are not impersonal numbers. They mirror real people's experiences in the health system and 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.[32] The majority of the data collection is based on a survey administered to the facility manager and includes direct observation of the availability and functioning of infrastructure, equipment, medicines, and other physical assets.

31 Kruk et al. (2018).

32 Details of sampling are discussed in Technical Appendix 1.

Structure of this chapter

Consistent with the SDI surveys' spirit, 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 patient caseload of these health care providers may limit their bandwidth to provide her appropriate care. Then it summarizes evidence on whether health workers are trained and skilled to correctly diagnose and treat the typical patient's condition, 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 better systems of monitoring and increased accountability, and, while some facilities have a large volume of patients, others are relatively under-utilized, 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 lower-level staff who form the frontline of care. Finally, some basic infrastructure, equipment, and medicines remain unavailable at many health facilities, representing 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 based on the findings across countries.

The findings presented in this chapter build on previously published work. A recent study using SDI health data assessed the quality of health care across Sub-Saharan Africa based on providers' clinical attendance, their knowledge of seven basic medical conditions, and the availability of key medicines, reporting country-level averages for key indicators and an overall estimate of care readiness.[33] The study's findings emphasized the need to strengthen health care providers' knowledge base, identifying this as the major limiting factor in further improving provision of care. The results showed how SDI tools and data can be used to diagnose key health-system challenges. In related work, Andrews et al. (2021) examine the correlates of these key indicators to assess differences in health facilities' performance within and between countries.[34] Their paper analyzes levels and trends in primary care service provision 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. This chapter is also designed to present results parallel to those in the education-focused chapter and allow for synthesis of findings across sectors.

33 Di Giorgio et al. (2020).

34 Andrews et al. (2020).

SAMPLE, METHODS, AND FRAMEWORK

The SDI health facility surveys have been implemented for over 10 years across 13 countries in Sub-Saharan Africa, and this chapter presents data combined from across this period.[35] In countries with multiple SDI surveys, only the most recent survey is included, since data from the same country over time were not fully comparable due to differences in sampling and measurement strategies.[36] The resulting dataset encompasses information on nine countries, covering 7,810 health facilities. This includes results from the following country surveys: Kenya (2018), Madagascar (2016), Mozambique (2014), Niger (2015), Nigeria[37] (2013), Sierra Leone (2018), Tanzania (2016), Togo (2013), and Uganda (2013).

In each country, the sample of surveyed facilities was selected based on a complete listing from the Ministry of Health of health facilities offering primary care services. This includes public facilities and those operated and administered by private entities or non-governmental organizations (NGOs). Given the surveys' focus on primary care, facilities at all levels of care, including 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 facility type. 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 Nigeria, where data were collected in 12 of 36 states due to logistical constraints. In Kenya data are representative not only at the national level but also at the county level.[38]

The SDI surveys purposefully include facilities operated by non-governmental entities. This is in line with the goal of providing an inclusive assessment of citizens' access to care. Since the private sector is responsible for a substantial portion 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, this includes private for-profit facilities (15.9 percent of sample), faith-based non-profit 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 based on the inverse probability of being sampled.[39] In all figures, countries are ordered by increasing GDP per capita to explore how performance on health service indicators may correlate with country income.[40]

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."[41] While the surveys did not comprehensively measure all aspects of quality encompassed by this definition, the data obtained cover eight key indicators: health care provider absence rate, 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, these 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 correctly diagnose and treat common conditions?

35 Five surveys were 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 (Guinea Bissau 2017, Cameroon 2019, and Malawi 2019).

36 See Chapter 4 for more on over-time comparability and results from SDI.

37 Note that the SDI survey on which the Nigeria write-up is based was carried out 8 years ago and in only 12 of the 36 states of Nigeria. Thus, it is not nationally representative and circumstance may have changed since then.

38 Details of the sample are presented in Table 1, with further detail in Appendix Table A1A and Technical Appendix 1.

39 Sample weights at the facility level are unavailable for Mozambique, so the unweighted results are reported. Provider-level weights are used for calculating the absence rate and for all clinical vignette related measures.

40 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 World Bank (2020a).

41 WHO, OECD and World Bank (2018).

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

These questions and the indicators embedded in them give a broad overview of the current state of the health system, with each indicator measuring a different aspect of quality of care.[42]

TABLE 1 SDI HEALTH SAMPLE SIZE BY COUNTRY

Country	Year	Sample		
		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

Notes: 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 sub-sample of up to 10 providers at each facility are selected for the clinical vignette interview, taken from the providers that report regularly providing outpatient care.

WILL HEALTH CARE PROVIDERS BE PRESENT IN THE HEALTH FACILITY?

After making the decision 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 those 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 care quality in low- and middle-income settings.[43] Despite recent gains, the problem persists in many countries. Better understanding of the underlying drivers and innovative solutions are needed.[44] The SDI surveys reveal that, on average, 6 out of 10 health care providers are present at the facilities surveyed.

Across all nine countries, 43 percent of providers are absent from their facility during an unannounced enumerator visit. This high absence rate includes absence for both authorized and unauthorized reasons (further analyzed below) and is consistent with rates observed across Bangladesh, Ecuador, India, Indonesia, Peru, and Uganda, where on average 35 percent of health workers were absent from health clinics during an unannounced visit.[45] SDI

42 Further details on the calculation of each indicator are available in Appendix Table A2.

43 Bamgboye and Adeleye (1992); Chaudhury and Hammer (2004); Belita, Mbindyo, and English (2013).

44 Tumlinson et al. (2019).

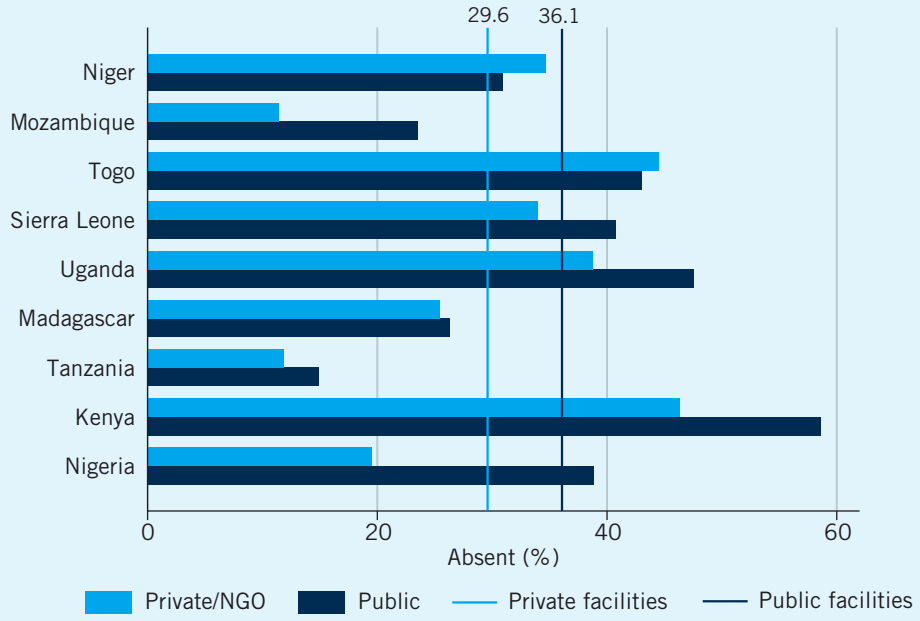
45 Chaudhury et al. (2006).

surveys include two visits, of which the second is unannounced, 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 over 50 percent in public facilities in Kenya to less than 20 percent in private facilities in Tanzania and Mozambique, as shown in **Figure 2**. Outside of 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 3** shows reasons for absence in public facilities, including training/meeting (21 percent), official mission (11 percent of absences), sick/maternity leave (11 percent), and other authorized absence (28 percent). Across all public facilities, only 4 percent of providers are absent without authorization. This was highest among doctors and clinical officers (COs) (5 percent) and nurses (4 percent), and lowest among other health workers (3 percent.[46] Overall, more than 90 percent of absences in public facilities are authorized. However, this is likely to be a conservative estimate, as 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 are more prevalent in public compared to private facilities in all nine countries. The higher rate of absence at public facilities may stem from differences in incentive structures, differences in management and supervision, or increased demands that take public health care providers outside of the facility. Some absences may be out of the control of facility administrators and there are national-level strategies that could be considered to keep staff on-site, such as limiting the frequency of government-mandated trainings or discouraging staff from maintaining secondary employment.

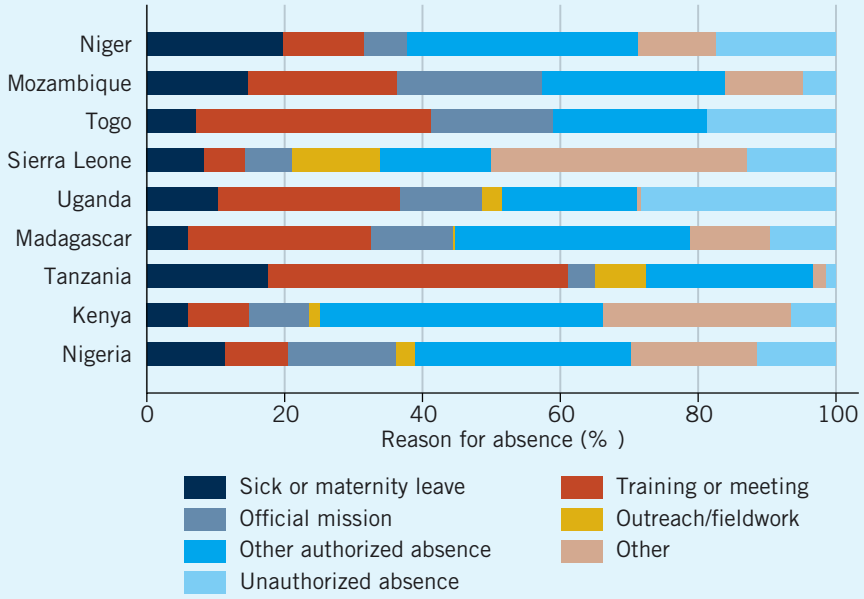
FIGURE 2 PROVIDER ABSENCE RATE BY FACILITY OWNERSHIP



Notes: 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.

46 The "other health workers" category includes a variety of positions which depend on the country. This 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.

FIGURE 3 **REASONS FOR HEALTH CARE PROVIDER ABSENCE IN PUBLIC FACILITIES**



Notes: This figure shows the reported reasons for absence among health care providers listed as absent in public facilities. The reasons for absence differed by country and have been grouped into major categories.

WILL HEALTH CARE PROVIDERS BE TOO BUSY WITH OTHER PATIENTS?

For primary care, when patients come to a health facility, they hope that there will be a medical professional 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 the patients who present in a day and that they are not overburdened with more patients than can be reasonably triaged. To estimate the care burden 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 the recognition that there is an uneven distribution of health workers, mismatched with health care needs. Countries with the lowest relative needs have the highest number of health workers, while 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. From a financial perspective, staff costs, in the form of salaries, allowances, and benefits, often account for over half of total health system cost.[47] 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. Caseload ranges from a low of 3 patients per provider per day in Nigeria to a high of 23 in Mozambique and Kenya. On average, providers in public facilities attend about 14 patients per day, while those in private attend about 12 patients per day. The pattern by rural/urban location varies, depending on the country, as shown in **Figure 4**.

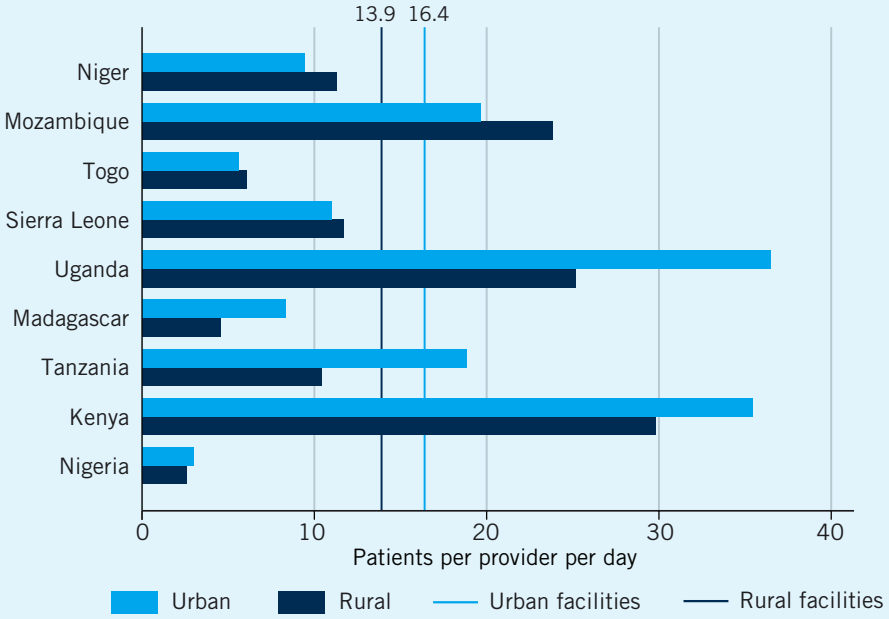
47 Kaplan et al. (2014)

While there is no "ideal" number of patients per provider per day,[48] the caseload estimates from the SDI surveys suggest a low caseload at many facilities. The share of facilities with low caseload (defined here as fewer than five patients per provider per day), ranges from 3 percent of public facilities in Uganda to 87 percent in Nigeria (**Figure 5**). Some of these differences may be due to sample composition differences between countries, with Nigeria including many lower-level facilities that appear to be less frequently visited. However, among public facilities overall, almost half (45 percent) have a caseload below five patients per provider per day. In contrast, in advanced economies like the United States, an average physician attends approximately 20.2 patients per day.[49]

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 similar overall outpatient volume as in the SDI survey and further details are shown in the box below. These findings are in line with recent analyses in Latin America, which looked at efficient allocation of care. In Ecuador, a World Bank report estimated that approximately one-third of primary care facilities had low case counts (fewer than 500 outpatient visits per month).[50] Similarly, in Peru, another analysis found that around 10 percent of health facilities had fewer than 30 total patient visits per day.[51] Although it is difficult to determine the optimal theoretical caseload, in both of these countries other similarly equipped facilities were able to see 3-4 times the volume of patients.

Caseload is only reflective of the direct patient interaction aspects of the health care providers' job, and their responsibilities may include other administrative and management tasks (such as maintaining medical records, managing procurement systems and supply stocks, and 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 seen in day-to-day operations. The

FIGURE 4 CASELOAD AMONG PUBLIC FACILITIES BY COUNTRY AND URBAN/RURAL FACILITY LOCATION



Notes: This figure shows the caseload by country and urban/rural difference. Caseload is based on the estimated number of outpatients per day (measured as the total outpatients in the prior three months), divided by the number of medical staff at the facility on a typical day.

48 Speakman (2016).

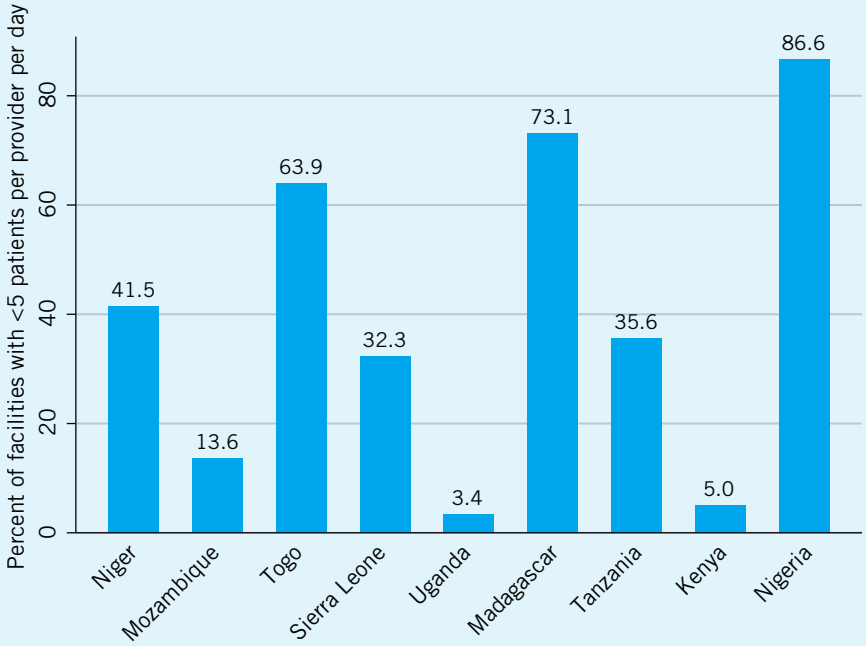
49 Hawkins (2012).

50 Vermeersch and Giovagnoli (2020).

51 World Bank (2020b).

low caseloads observed therefore do not necessarily suggest that health care providers have ample time to see more patients, as their time may be stretched to accommodate other activities. However, it does suggest that some commonly reported problems, such as overcrowding and long wait times, may be due to other factors, such as poor facility management, large administrative burdens, or an uneven distribution of patients throughout the day, rather than a lack of sufficient staff.

FIGURE 5 SHARE OF PUBLIC FACILITIES WITH FEWER THAN FIVE PATIENT VISITS PER PROVIDER PER DAY, BY COUNTRY



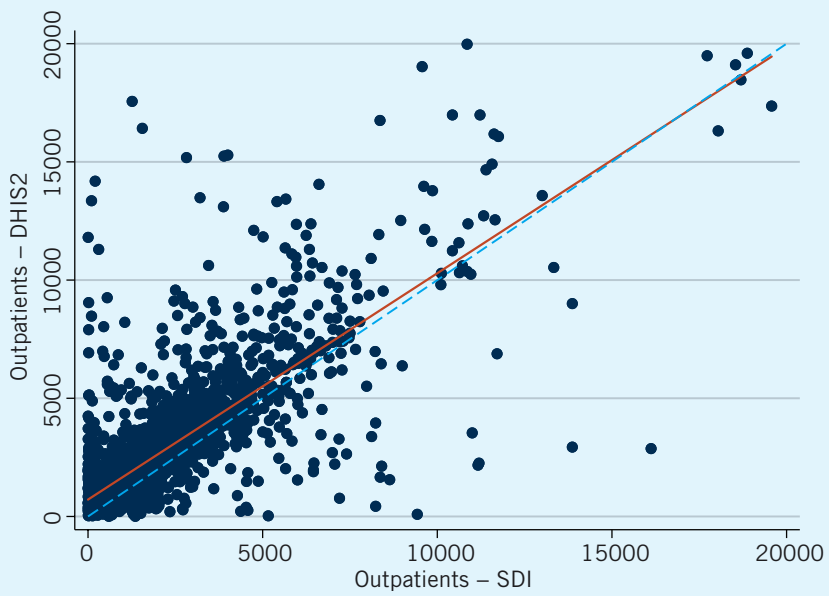
Notes: This figure shows the total rate of absence during a second, unannounced visit. The denominator is up to 10 randomly selected health care providers listed on the roster at each facility.

BOX 2: TRIANGULATING SDI SURVEY FINDINGS WITH ADMINISTRATIVE DATA

Caseload estimates in the SDI surveys are slightly lower than expected, at odds with researchers' initial hypothesis and the literature on human resource shortages in health care.* 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 health management information systems (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 using facility name and district. Data were missing or zero in DHIS2 for 439 of 3,034 facilities (14.5 percent).

Figure 6 depicts a simple scatter plot of the sum of outpatients as recorded by DHIS2 versus the SDI estimates. The overall correlation is 0.79. While 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 6 COMPARISON OF OUTPATIENT NUMBERS IN DHIS2 VERSUS SDI SURVEYS, WITH LINE OF EQUIVALENCY (RED) AND BEST FIT (GREEN), KENYA, 2018



Notes: Each dot in this figure represents one facility, with the DHIS2 reported outpatients on the y-axis and the SDI reported outpatients on the x-axis. The orange line is a line of equivalency (x = y) and the blue line is the line of best fit.

* Liu et al. (2016).

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 providers accurately diagnose and appropriately treat illnesses, is not a given. SDI data can shed light on the facets of the patient experience that have to do with health care providers' clinical skills. 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 an increasing recognition that good health outcomes depend not just on patients' access to care but 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 utilization.[52] The SDI survey includes clinical vignettes that are administered to health care providers. This is an innovative addition that measures the quality of clinical care, rather than the inputs-focused perspective that many earlier surveys had taken.[53] Clinical vignettes may be less reliable for assessing quality of care than other methods, such as the use of standardized patients, but 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."[54]

In the SDI surveys considered in this booklet, health care providers are tested on five core vignettes: childhood diarrhea with dehydration, childhood pneumonia, adult tuberculosis (TB), adult diabetes mellitus, and childhood malaria with anemia. Additionally, countries may choose to add specific vignettes and occasionally remove vignettes (as in Kenya, where the malaria and anemia vignette was not administered). These vignettes represent common clinical cases that a health care provider would face in the low- and middle-income country context. The diseases included are high-burden conditions, making up 30 percent of all-age disability-adjusted life years (DALYs) in Sub-Saharan Africa.[55] Each provider is scored on the percentage of vignettes for which they provide the correct diagnosis and treatment.[56] Multivariate regressions to test the relationship between provider-level variables are described here but presented fully in a previous paper.[57] Further details on the vignettes are available in Appendix Tables A3 and A4.

Clinical vignettes are useful to measure health care provider diagnostic and treatment accuracy. They also provide valuable information on their adherence to clinical protocols. The vignettes contain country-adapted information to simulate a full consultation, including recommended patient history questions, a 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 3.5 percent of health care providers accurately diagnose diarrhea with severe dehydration. To assess the severity of the case, the Guidelines for Integrated Management of Childhood Illnesses (IMCI) recommend administering a skinfold test, checking for lethargy, checking whether the child is able to drink, and checking for agitated/irritable behavior. IMCI requires two positive danger signs to denote a case as severe, but only 24 percent of providers asked two questions. 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 our understanding of provider performance and offer insights for policy or investment actions.

Figure 7 shows the confidence interval estimates for diagnostic accuracy by provider cadre and country. Overall mean diagnostic accuracy (combining the results from all vignettes) varies by country, from a high of 69 percent in Tanzania to a low of 40 percent

52 Rao and Sheffel (2018); Escamilla et al. (2018).

53 Das and Leonard (2006).

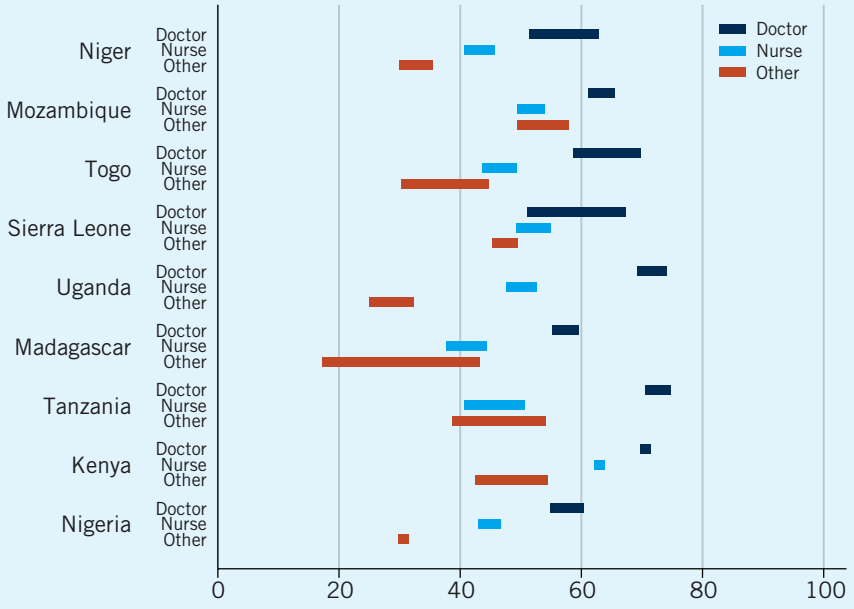
54 Peabody et al. (2000).

55 Diarrhea at 7.6 percent of DALYs, 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).

56 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).

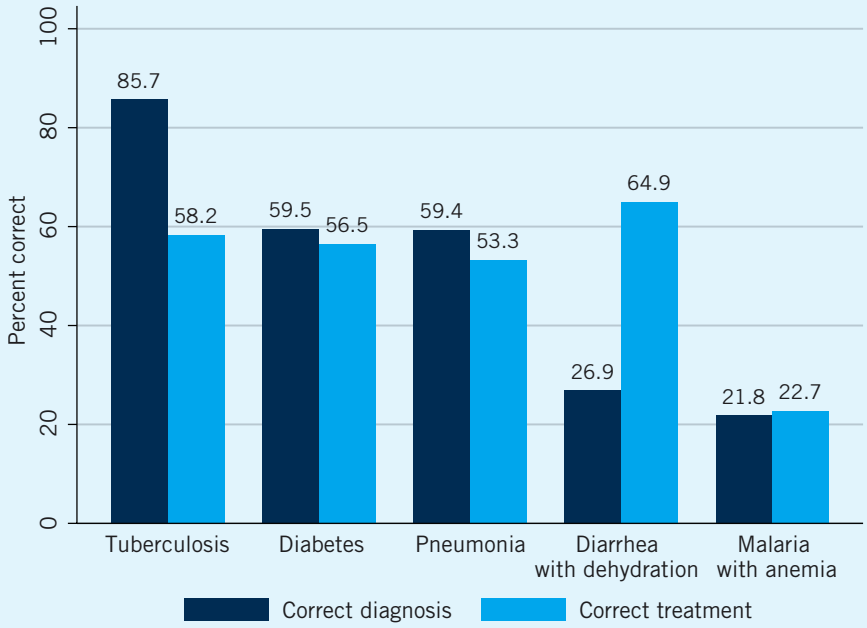
57 This section includes results from multivariate regressions presented in Andrews et al. (2021). Multivariate regressions for diagnostic and treatment accuracy include facility and provider level controls. The facility level controls are ownership (public/private), location (urban/rural), facility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

FIGURE 7 DIAGNOSTIC ACCURACY, BY COUNTRY AND HEALTH CARE PROVIDER CADRE



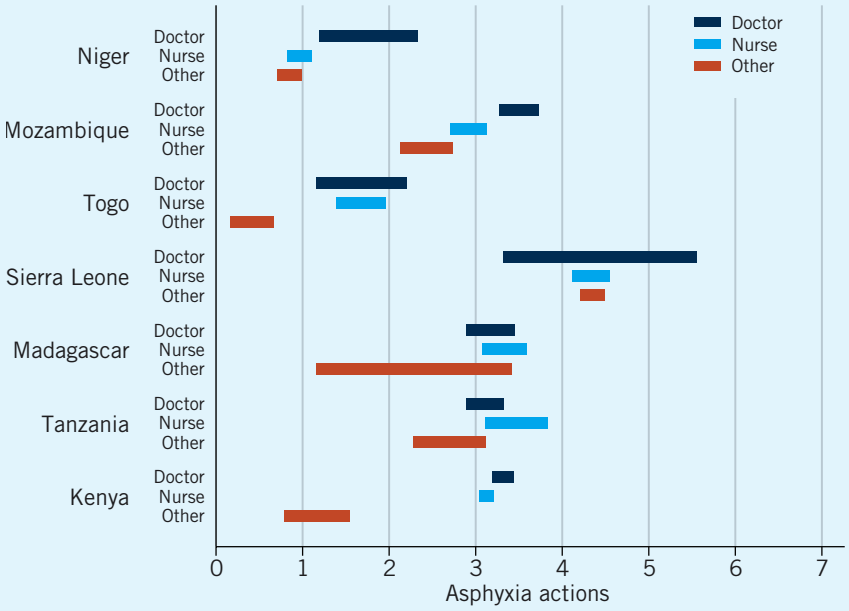
Notes: Diagnostic accuracy is combined across the five core vignettes, shown in Figure 8. The ranges shown represent 95% confidence intervals.

FIGURE 8 HEALTH CARE PROVIDER ACCURACY OF DIAGNOSIS AND TREATMENT, BY DISEASE, IN PUBLIC FACILITIES



Notes: 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

FIGURE 9 ACTIONS TAKEN FOR NEONATAL ASPHYXIA IN PUBLIC FACILITIES, BY COUNTRY AND HEALTH CARE PROVIDER CADRE*



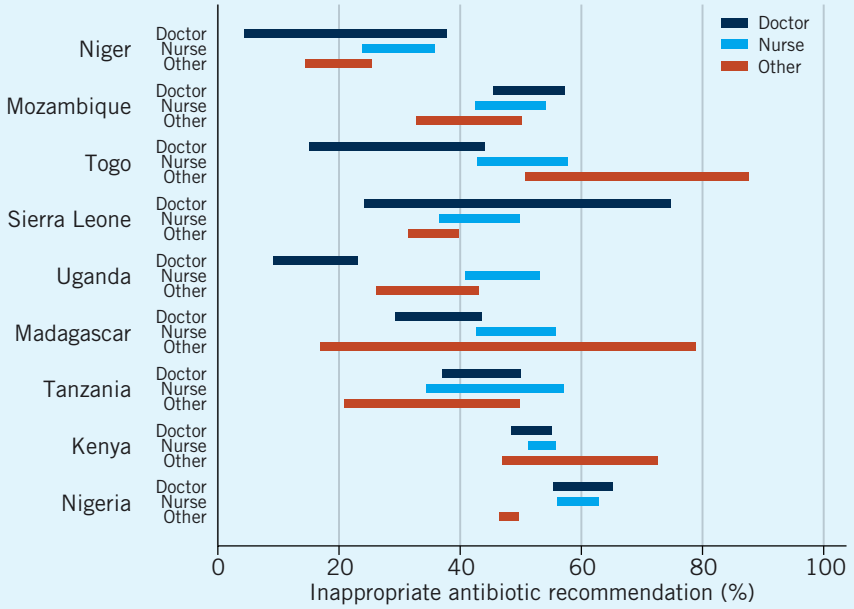
Notes: Health care providers are graded based on the number of actions that they mention (out of 7 possible actions) for responding to a case of neonatal asphyxia. The ranges shown represent 95% confidence intervals for the average number of actions recommended.

* Nigeria and Uganda are omitted due to incomparability of the standards used in assessment.

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 between the cadres 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 40-49 year olds have slightly higher diagnostic accuracy than younger or older colleagues. Perhaps not unexpectedly, providers with secondary and post-secondary education perform significantly better than those with only primary education. There is not a significant difference between 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. Since lower-level providers will likely continue to provide the bulk of diagnoses at 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 8**. 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% of cases for all diseases, except malaria with anemia which has a correct treatment rate of 22.7%. Doctors and clinical officers are more likely to perform better in treatment accuracy than nurses and other staff, providers with post-secondary education performing better and male providers performing better overall. Provider treatment accuracy is higher at hospitals than other facility types, higher at rural than urban facilities, and higher at public than private facilities. Treatment accuracy is higher than diagnosis accuracy for the two dual diagnosis conditions because providers sometimes failed to diagnose the comorbidity (e.g. diagnosing malaria, but not anemia) but offered treatment that was satisfactory for both conditions (e.g. artemisinin combination

FIGURE 10 INAPPROPRIATE ANTIBIOTIC RECOMMENDATION IN PUBLIC FACILITIES, BY COUNTRY AND HEALTH CARE PROVIDER CADRE



Notes: Inappropriate antibiotic prescription is defined as providers that prescribe an antibiotic during the tuberculosis vignettes (aside from the antibiotics recommended as part of the tuberculosis regimen) or any antibiotics for the diarrhea vignettes (for which antibiotics are not indicated given the patient examination). The ranges shown represent 95% confidence intervals for the percent of health care providers giving an inappropriate antibiotic prescription.

therapy and iron supplements).[58] These results suggest that patients are likely to receive insufficient treatment, particularly where multiple conditions are present, and that this problem is more common when treatment is provided by lower-level staff.

To offer information about management of urgent maternal and child health conditions, SDI survey includes two additional vignettes, one on neonatal asphyxia and one on post-partum hemorrhage. Both conditions have a clear set of guidelines and recommended actions that the providers should undertake in order to reduce the risk of maternal and neonatal morbidity and mortality. For the purposes of scoring, providers are assessed on seven actions for neonatal asphyxia (calling for help, placing the baby in a neutral position, checking the baby's heartrate, checking the baby's breathing, drying the baby, keeping the baby warm and initiating resuscitation with a bag/mask) and five actions for post-partum hemorrhage (determine the cause of the post-partum hemorrhage, provide bimanual uterine massage, place a foley catheter, run an IV, and provide oxytocin or similar drugs). Providers are scored on the number of these correct actions that they propose. Results for neonatal asphyxia by provider cadre are shown in **Figure 9**. Assessing providers on these two measures, nurses score as well as doctors/COs on the asphyxia vignette and almost as well on the PPH 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, though it is encouraging that nurses score comparably to doctors on this measure.

Finally, the SDI survey also collects information on inappropriate antibiotic use, an ongoing global challenge which can contribute to the rise of antibiotic resistance.[59] In the SDI surveys, inappropriate

58 In the diarrhea and dehydration vignette, the child presents as a case of diarrhea but displays multiple warning signs for severe dehydration. According to Integrated Management of Childhood Illness (IMCI) guidelines, severe dehydration necessitates rehydration with an IV or NG tube. However, 86 percent of providers simply prescribe oral rehydration salts (ORS), and 45 percent prescribe ORS plus zinc. ORS + zinc is scored as appropriate treatment, since the child was 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.

59 Laxminarayan et al. (2013).

antibiotic prescription is defined as providers that prescribe an antibiotic during the tuberculosis vignettes (aside from the antibiotics recommended as part of the tuberculosis regimen) or any antibiotics for the diarrhea vignettes (for which antibiotics are not indicated given the patient examination). Inappropriate antibiotic usage is calculated as the percentage of health care providers that inappropriately prescribe antibiotics among all health care providers given the clinical vignettes. Results for inappropriate antibiotic usage are shown in **Figure 10** and indicate that nearly half of health care providers prescribe an antibiotic in cases where it is not recommended. The results by cadre show variation in prescription patterns, with doctors more likely to inappropriately prescribe antibiotics in Mozambique, Sierra Leone, and Nigeria, while nurses or other medical staff are more likely to inappropriately prescribe antibiotics 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 is higher at hospitals compared to health centers or health posts. Health care providers over the age of 50 are less likely to inappropriately prescribe antibiotics than younger providers, and males are more likely to inappropriately prescribe antibiotics. In some cases, antibiotics are inappropriately ordered in addition to correct treatment, which still suggests widespread over-prescription of antibiotics. The variation suggests that health care providers with more education or training are not always more likely to make better prescription decisions. While clear guidelines and oversight of antibiotic usage could help reduce over-prescription, the growing recent literature points at knowledge gaps, misaligned incentives, and patients' own demand as possible causes of overprescription.[60]

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, high-quality care may be constrained by availability of key inputs such as basic infrastructure, equipment, clean water, supplies, and medicines. 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.[61] The availability of these physical resources does not guarantee system competence, but any gap in 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.[62]

Infrastructure availability

Basic infrastructure availability—defined as the availability of an improved water source, improved toilet, and electricity[63]—varies from a country average of 77.2 percent in Kenya to 21.0 percent in Niger. Infrastructure availability is significantly higher in urban areas, driven partly by the higher rates of electricity availability. 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 11 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 Sierra Leone and Madagascar, compared to richer countries like Tanzania and Kenya. **Figure 12** 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 fewer facilities lack more than one basic infrastructure item, whereas in Niger a greater

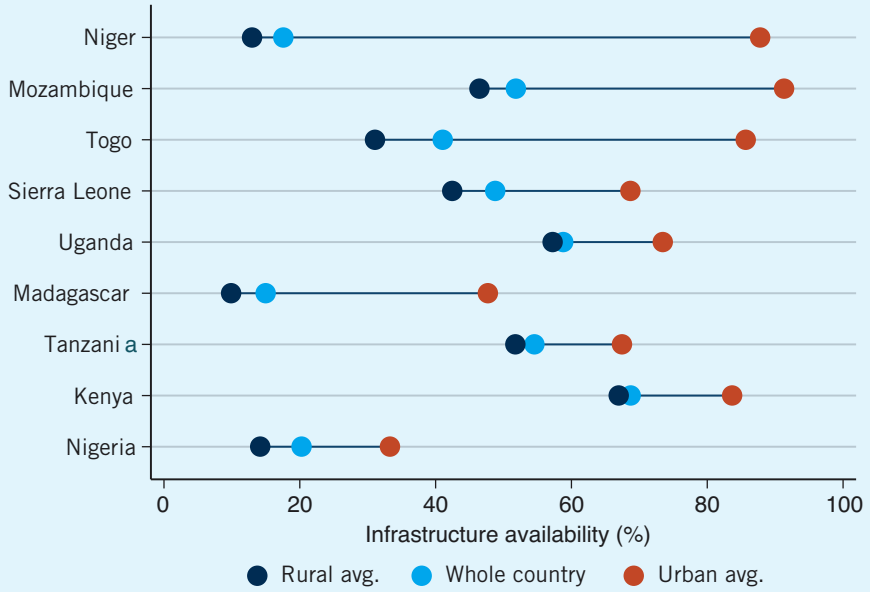
60 Lopez, C., A. Sautmann, and S. Schaner (2021).

61 UNICEF (2019).

62 Leslie, Sun, and Kruk (2017).

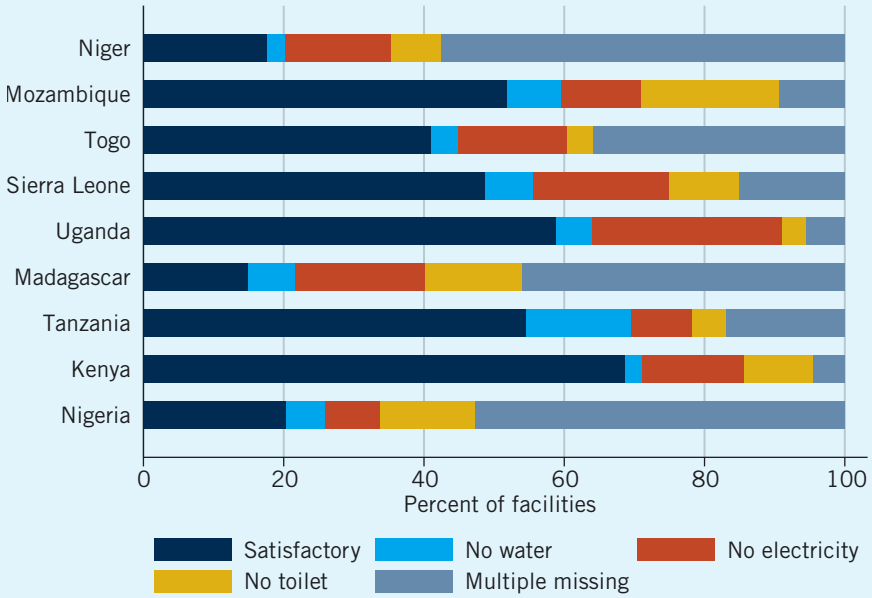
63 Definitions for improved water source and improved toilet are detailed in Appendix Table A2.

FIGURE 11 AVAILABILITY OF BASIC INFRASTRUCTURE IN PUBLIC HEALTH FACILITIES, BY COUNTRY AND URBAN/RURAL LOCATION



Notes: This figure shows the percentage of facilities with all basic infrastructure available (improved water, improved sanitation, and electricity). Dark blue dots represent the average score at rural facilities, orange dots represent the average score at urban facilities, and blue indicates the whole-country estimate.

FIGURE 12 BREAKDOWN OF INFRASTRUCTURE AVAILABILITY (IMPROVED WATER SOURCE, IMPROVED SANITATION, ELECTRICITY) IN PUBLIC HEALTH FACILITIES, BY COUNTRY



Notes: This figure shows the breakdown of infrastructure not available in public facilities by country. Multiple missing refers to facilities that have more than one piece of infrastructure unavailable.

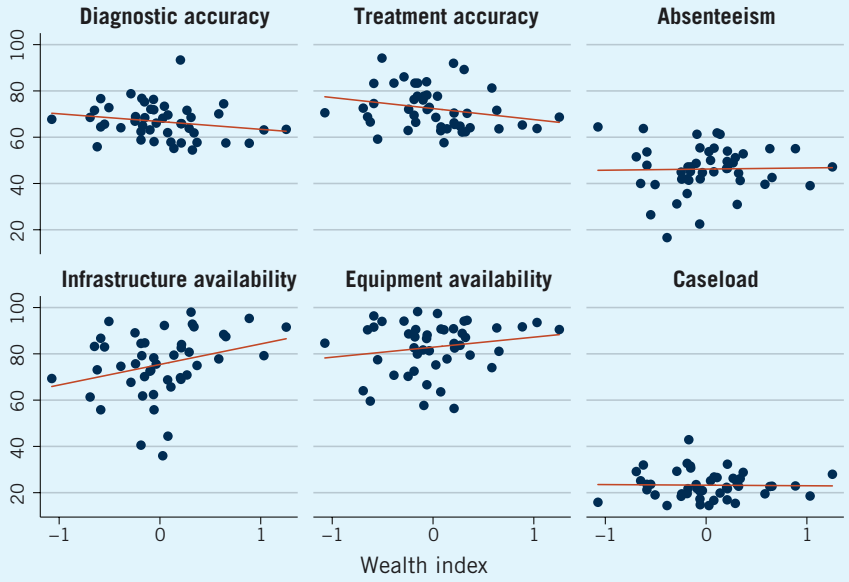
BOX 3: FACILITY CHARACTERISTICS AND WEALTH: EVIDENCE FROM KENYA

Policy makers have pursued equity in health care provision for many decades, acknowledging that health systems should not just aim to maximize coverage but also to ensure that care is available for those who need it most.* By measuring 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 Kenya SDI survey data to the country's 2014 Demographic and Health Survey (DHS) data allows for examination of the relationship between facility-level characteristics and household wealth.

The DHS 2014 includes 36,430 households, each with an estimated wealth index constructed using a 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 13**). The only significant relationship was for health facility infrastructure availability, which was lower in counties with a lower wealth index. However, this relationship may be driven by the greater share of poor households in rural areas, as the relationship does not remain significant when controlling for the share of DHS households that are rural.

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

FIGURE 13 DHS WEALTH INDEX AND FACILITY-LEVEL CHARACTERISTICS IN KENYAN COUNTIES



Notes: Each dot represents a Kenyan county, with average county wealth index score on the x-axis. Results for six major SDI indicators are shown, with the county-level average score on the y-axis.

* WHO (2008).
** McCollum et al. (2018).

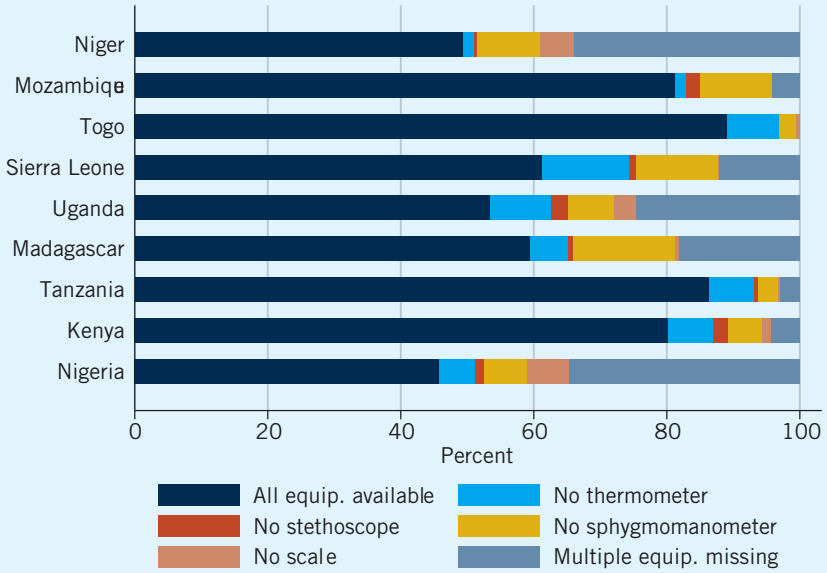
proportion of facilities have multiple deficits. Similarly, the SDI results reveal that lack of access to electricity is a problem across all countries, but that this deficit is most pronounced in Uganda, whereas lack of improved toilet and improved water are more common in Nigeria and Tanzania, respectively. **Figure 12** 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 in infrastructure availability within and between countries, what are the strongest predictors of basic infrastructure availability? Multivariate regressions show that country, health facility level, facility ownership, and urban location are all important determinants of basic infrastructure availability. The urban/rural gap was 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. Infrastructure availability is lower in Kenyan counties with a lower household wealth index, as shown in **Box 3**. 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 the tools and medicines needed for proper diagnosis and treatment. The list of equipment and medicines necessary can vary depending on the facility level and type of care provided, but the SDI survey focuses on basic items. For equipment, that means tools such as a thermometer, stethoscope, blood pressure cuff, and weighing scale. These represent the bare minimum items that should be present in all facilities that are offering preventive and curative health care services. For medicines, the survey focuses on 14 common medicines, a subset of WHO's Model List of Essential Medicines, and availability is calculated as the percentage of those 14 medicines which are in-stock and unexpired on the day of the facility visit.[64] Are all facilities equipped with these essential pieces of equipment and common medicines?

FIGURE 14 BREAKDOWN OF EQUIPMENT AVAILABILITY IN PUBLIC FACILITIES, BY COUNTRY



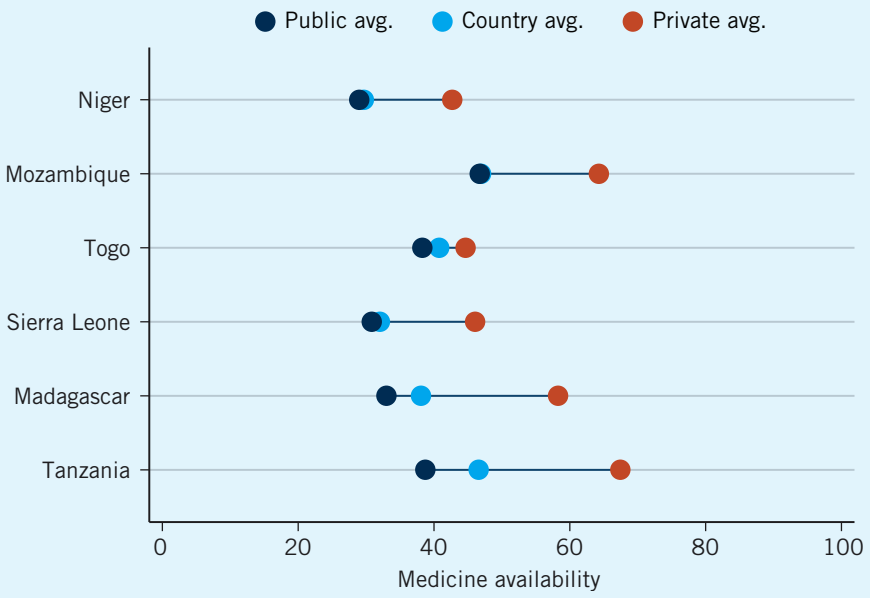
Notes: This figure shows the breakdown of missing equipment in public facilities by country. Multiple equip. missing refers to facilities which have more than one piece of equipment unavailable.

64 WHO (2017).

The SDI surveys indicate that 67 percent of public and 79 percent of private facilities have all four pieces of equipment (i.e., thermometer, stethoscope, blood pressure cuff, and weighing scale). Availability of basic equipment is better in public facilities 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) compared to rural facilities (66 percent) has all four pieces of equipment. Although highly variable across the nine countries, there is no clear relationship between availability of these basic tools and country-level living standards as proxied by average GDP per capita. **Figure 14** shows the key pieces of equipment missing in different countries. Often, facilities are missing only one of the four necessary pieces of equipment, but over 30 percent of facilities in Nigeria and Niger lack multiple items. Personal protective equipment is another important set of equipment for health facilities and is profiled in **Box 4** for Sierra Leone.

Medicine availability is notably lower at health posts (32 percent) and health clinics (46 percent), compared to 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 facilities (36 percent) compared to private facilities (54 percent), as shown in **Figure 15**. The difference between private and public facilities varies by country, from a 6-percentage-point difference in Togo to a 29-percentage-point difference in Tanzania. Medicines are about equally available in rural and urban areas, suggesting somehow surprisingly, that the location of facilities does not influence their supply stocks. 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 available. Of note, some medicines, such as oxytocin, require refrigeration to prevent spoilage, and only two-thirds of facilities (66 percent) report having a functional refrigerator.

FIGURE 15 PERCENTAGE OF THE 14 BASIC MEDICINES AVAILABLE IN FACILITIES, BY COUNTRY AND OWNERSHIP TYPE



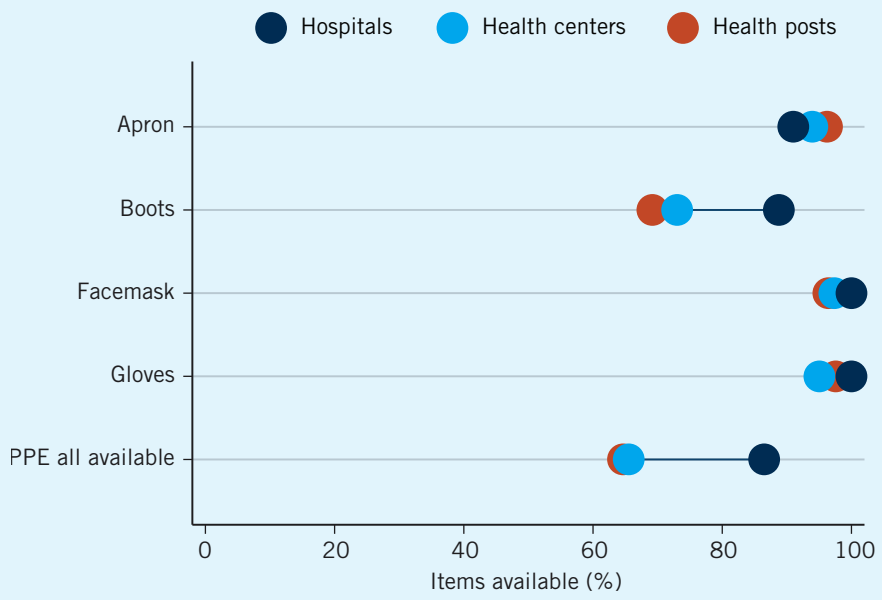
Notes: This figure shows the percent of medicines available by country and facility ownership. Medicine availability is calculated as the share of 14 essential medicines available at a facility. The dark blue dots represent the average score at rural facilities, orange dots represent the average score at urban facilities, and blue indicates the whole-country estimate.

BOX 4: PANDEMIC PREPAREDNESS IN SIERRA LEONE

The COVID-19 pandemic has called increased attention to systems' ability to protect health workers from acutely infectious disease. Personal protective equipment (PPE) is used to ensure that health care workers are protected when serving high-risk patients. The Sierra Leone SDI survey took place in 2018, following the 2013-2016 West African Ebola epidemic, and collected data on availability of disposable gloves, apron, face mask, and protective boots, though information on protective eyewear and face shields was not available. Encouragingly, disposable gloves, face mask, and apron 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 together in 66 percent of facilities.

Figure 16 shows the availability of these items for hospitals and non-hospitals (health clinics and health posts). The high rates of PPE ownership may be partly attributable to the Ebola response. Importantly, the SDI surveys (like other large-scale health facility 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 infeasible; however, it also means that the estimates of availability described here do not necessarily reflect adequate supply. Looking forward, SDI surveys will include more information on pandemic preparedness (see Chapter 4).

FIGURE 16 AVAILABILITY OF PERSONAL PROTECTIVE EQUIPMENT IN HEALTH FACILITIES IN SIERRA LEONE, BY FACILITY TYPE



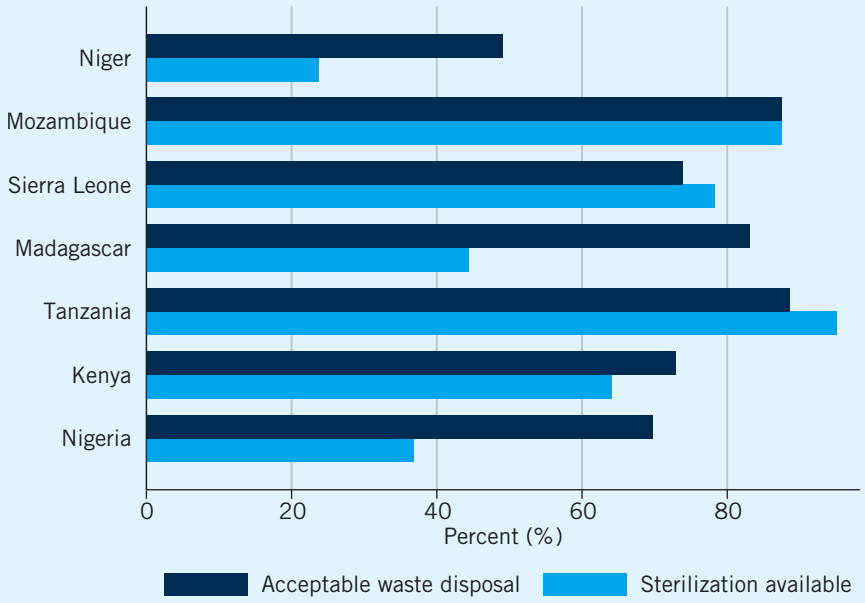
Notes: This figure shows the percentage of facilities with different pieces of personal protective equipment (PPE) available in Sierra Leone, with the dots showing discrepancies by facility type.

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 infection prevention and control (IPC) are always important but especially so during a pandemic. Hand hygiene and appropriate waste segregation and disposal, for instance, are necessary for both patient and health care worker safety. The necessity of appropriate water, sanitation, and hygiene (WaSH) techniques was demonstrated during the 2014 Ebola epidemic in Sierra Leone, where the rate of infection among health care workers was several times higher than among the general population, partly owing to lack of robust IPC measures in place in health facilities.[65] 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 health clinics (77 percent) or health posts (76 percent) and are also higher in urban areas and at private facilities. Similarly, equipment for sterilization is more widely available at hospitals (91 percent) than 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 scored lowest for both indicators, and Tanzania achieved the best scores. Countries tended to score higher on waste disposal than sterilization, as shown in **Figure 17**.

FIGURE 17 AVAILABILITY OF ACCEPTABLE WASTE DISPOSAL AND STERILIZATION IN PUBLIC FACILITIES, BY COUNTRY



Notes: This figure shows the percentage of facilities with acceptable waste disposal and sterilization available. The definition for both was based on WHO standards.*

* WHO (2014)

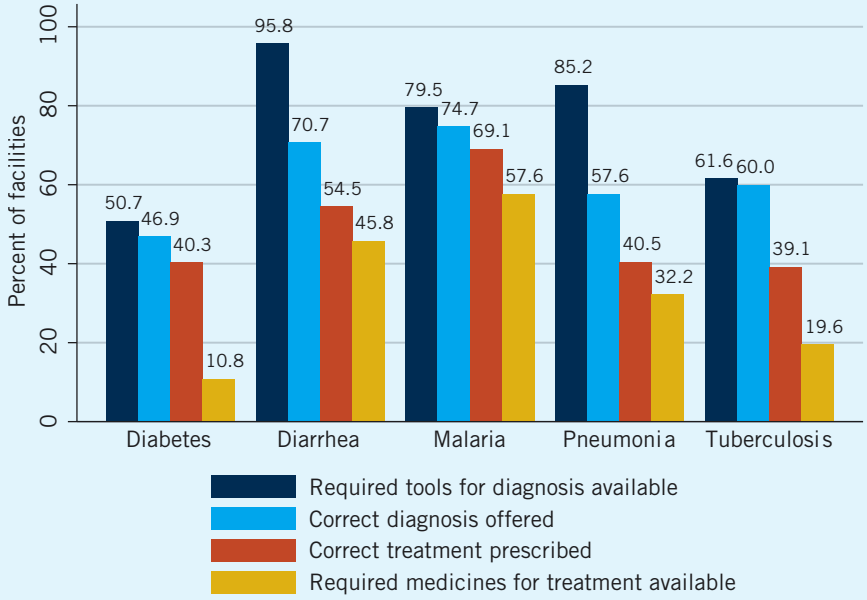
65 Kilmarx et al. (2014).

CAN AN AVERAGE CITIZEN'S EXPERIENCE WITH THE HEALTH SYSTEM BE IMPROVED?

The analysis presented above describes individual facets of service delivery. However, the experience of a typical patient is the result of the 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 care readiness. This analysis is intended to demonstrate how the wealth of information in SDI surveys can be recombined in novel ways to show system limitations.

Figure 18 depicts the availability of tools for diagnosis, the knowledge of health care providers in diagnosis and treatment, and the availability of medicines for treatment, for 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 each condition and correct diagnosis and treatment refers 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 the inputs in the step prior, with the final bar representing the overall likelihood of receiving all 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 required medicine for treatment is artemisinin combination therapy and there needs to be at least one provider in the facility who can offer accurate diagnosis and treatment for

FIGURE 18 AVAILABILITY OF KEY INPUTS FOR DIAGNOSIS AND TREATMENT OF FIVE COMMON CONDITIONS



Notes: Equipment for diagnosis was defined as: a scale for diarrhea, a thermometer and rapid diagnostic test (RDT) for malaria and a stethoscope and thermometer for pneumonia. Appropriate equipment for diabetes and tuberculosis was not included in the SDI surveys and health care providers' mention of equipment in the vignettes is used as a proxy. For diabetes, appropriate equipment was assumed to be available if at least one health care provider at the facility ordered a fasting blood sugar, random blood sugar, HbA1c, or urinalysis test. For tuberculosis, appropriate equipment was assumed to be available if at least one health care provider ordered a chest x-ray, sputum microscopy or ESR/CRP test. In general, these pieces of equipment are not intended to be exhaustive but are a selection of relevant equipment. Medicines for treatment were defined as: hypoglycemics or insulin for diabetes, ORS for diarrhea, artemisinin combination therapy (ACTs) 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/treatment and/or malaria medicines. Kenya, Nigeria, Togo and Uganda are excluded from the tuberculosis section because they did not have information on tuberculosis combination therapy.

malaria on the clinical vignettes. Although most facilities had these individual components, only a little over half of facilities have all the necessary components in combination and can therefore be considered prepared to treat a malaria case.

Care readiness 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 was the availability of the 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 begin to address this problem or to provide appropriate medicines.[66] For the other conditions, no single factor emerges as dominant, but 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 facility surveys give insight into ordinary people's experience of primary health care 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' 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 be unavailable. The SDI surveys show that large numbers of health facilities still lack the basic infrastructure (electricity, water, and 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 primary care service delivery. The service delivery agenda moving forward should entail both an expansion and reorganization of care to simultaneously manage immediate risk and address long-term challenges.[67] The results from the analysis presented herein 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 private ones, suggesting that practitioners in public facilities may have increased demands that take them outside of the facility, including attending trainings, outreach, and 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. This is an important entry point for policy, as differences in provider absence rates between public and private facilities have the potential to further increase health disparities.

Re-balance caseloads and resources system wide. As noted, low caseloads might raise concerns about effective allocation of human resources in health care delivery, but they also point to the success of the global effort to bring health facilities and health care

66 Manne-Goehler et al. (2019); Raviglione (2012).

67 World Bank (2020e).

providers closer to people, especially in rural and remote areas.[68] While expanded geographical access to care is important, especially from an equity perspective, low caseloads in primary care facilities across countries provide further evidence to consider reorganization of services within the existing health system to enhance efficiency without compromising on equitable access, as proposed by the Lancet Global Health Commission on High Quality Health Systems.[69] Quality-focused service delivery redesign would entail providing treatment of chronic and stable conditions, preventive care, low acuity and algorithmic services, and palliative care at the primary level, while more complex or rare conditions would be managed in tertiary or specialized care centers.

Reinforce competencies among non-physician 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 successfully diagnose and treat. Diagnostic and treatment accuracy are higher at hospitals than health centers or health posts. Since the majority of initial patient presentation is likely to be at lower-level facilities, these results suggest that strengthening the capacity of workers at lower-level institutions is needed to enable them to act as the front line of primary care. Nurses and lower-level medical providers comprise the majority of the health workforce and need to be relied upon for patient care. Yet, they perform significantly worse than doctors on diagnostic and treatment accuracy as measured by the SDI vignettes. The evidence suggests that more attention should be placed on the skills and training of the health workforce to ensure patients receive quality care, regardless of the type of provider they see. Improving these competencies likely requires attention to the quality of clinical education and improvements to the existing curriculum, beyond the current standards for in-service training. On the positive side, results on 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 persists even in urban areas, suggesting that it is not driven by the inaccessibility or remoteness of facilities. Further investigation is required to better understand the potential incentive structures, efficiencies, and management practices that may be driving better supply chain management performance 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. While 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 facility staff and patients are at increased risk of infection and associated illness.[70] 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 storage of vaccines or other medicines.[71] Scaling up solar power infrastructure at health facilities may help address some of these cold chain challenges. The capacity for cold chain maintenance is taking on added importance now, as the global community begins provision of COVID-19 vaccines.[72] Given that rural health facilities may also be more likely to provide care to lower-income members of the population, poorer-quality facility infrastructure may also contribute to increased 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 a number of 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 health system redesign efforts. 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.

68 WHO and World Bank (2017).

69 Kruk et al. (2018).

70 WHO and UNICEF (2019); Sharma et al. (2020).

71 Adair-Rohani et al. (2013).

72 Fischetti (2020).

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):249-261. https://doi.org/10.9745/GHSP-D-13-00037

Andrews, K., Conner, R., Gatti, R., Sharma, J. 2021. "The Realities of Primary Care: Variation in Quality of Care Across Nine Countries in Sub-Saharan Africa." Policy Research Working Paper; No. 9607. World Bank, Washington, DC.

Bamgboye, E.A., and A.I. Adeleye. 1992. "Sickness absenteeism in a Nigerian teaching hospital." *East African Medical Journal* 1992 Aug;69(8):450-5.

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*. 2013 Jul 1:11-34. doi: 10.1186/1478-4491-11-34.

Chaudhury, N., and J. Hammer. 2004. "Ghost Doctors: Absenteeism in Rural Bangladeshi Health Facilities", *The World Bank Economic Review*, Volume 18, Issue 3, 2004, Pages 423–441

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 No. 9440, World Bank, Washington, DC.

Endler, M., T. Al- Haidari, C. Benedetto, S. Chowdhury, J. Christilaw, F. El Kak, D. Galimberti, C. Garcia- Moreno, M. Gutierrez, S. Ibrahim, and S. Kumari. 2020. "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.* 2020 Nov 11. 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: 1–12. 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/

Kaplan, R.S., M. Witkowski, M. Abbott, A.B. Guzman, L.D. Higgins, J.G. Meara, E. Padden, A.S. Shah, P. Waters, M. Weidemeier, and S. Wertheimer. 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, D.E. Sugerman, J.S. Bresee, J. Mermin, and J. McAuley. 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, P. Barker, B. Daelmans, S.V. Doubova, M. English, E.G. Elorrio, F. Guanais, O. Gureje, L.R. Hirschhorn, L. Jiang, E. Kelley, E.T. Lemango, J. Liljestrand, A. Malata, … and M. Pate. 2018. "High-quality health systems in the Sustainable Development Goals era: time for a revolution." *The 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. 2018b. "Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries." *The Lancet* 2018 Nov 17;392(10160): 2203-2212.

Laxminarayan, R., A. Duse, C. Wattal, A.K. Zaidi, H.F. Wertheim, N. Sumpradit, E. Vlieghe, G.L. Hara, I.M. Gould, H. Goossens, and C. Greko. 2013. "Antibiotic resistance—the need for global solutions." *The Lancet Infectious Diseases* 13(12): 1057-1098.

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.

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

Liu, J., Y. Goryakin, A. Maeda, T. Bruckner, and R. Scheffler. 2016. "Global health workforce labor market projections for 2030." World Bank Policy Research Working Paper No. 7790, World Bank, Washington, DC.

Manne-Goehler, J., P. Geldsetzer, K. Agoudavi, G. Andall-Brereton, K.K. Aryal, B.W. Bicaba, P. Bovet, G. Brian, M. Dorobantu, G. Gathecha, and M. Singh Gurung. 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): 729–742. 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.

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): 1715-1722.

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

Physician's Foundation. 2012. "A Survey of America's Physicians: Practice Patterns and Perspectives." Merritt Hawkins.

Rao, K.D., and A. Sheffel. 2018. "Quality of clinical care and bypassing of primary health centers in India." *Social Science and Medicine* 207: 80–8.

Raviglione, M., B. Marais, K. Floyd, K. Lönnroth, H. Getahun, G.B. Migliori, A.D. Harries, P. Nunn, C. Lienhardt, S. Graham, and J. Chakaya. 2012. "Scaling up interventions to achieve global tuberculosis control: progress and new developments." *The Lancet* 379(9829): 1902-1913.

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 n maternal and child mortality in low- and middle- income countries." *The Lancet Global Health* 2020 Jul;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 No. 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. https://www.wolterskluwer.com/en/expert-insights/effective-interprofessional-education?utm&utm_landingPage=https://www.wolterskluwer.com/en/expert-insights/effective-interprofessional-education&utm_prevPage=https://www.google.com/

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), pp.1-8.

UN General Assembly. 2015. "Transforming our World: the 2030 Agenda for Sustainable Development." 21 October 2015 A/RES/70/1. https://www.refworld.org/docid/57b6e3e44.html

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. World Health Organization, Geneva.

United Nations. 2011. "Millennium Development Goals Report 2011," June 2011, ISBN 978-92-1-101244-6, available at: https://www.refworld.org/docid/4e42118b2.html

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

WHO. 2013a. *World Health Report 2013: Research for universal health coverage*. WHO, Geneva.

WHO. 2013b. "Arguing for universal health coverage." WHO, Geneva. https://apps.who.int/iris/handle/10665/204355

WHO. 2014. "Safe management of wastes from health-care activities." WHO, Geneva.

WHO. 2019. "Model List of Essential Medicines." WHO, Geneva.

WHO. 2020a. "Continuity of essential health services: Facility assessment tool." WHO, Geneva.

WHO. 2020b. "Role of primary care in the COVID-19 response." Manila: WHO Regional Office for the Western Pacific. https://apps.who.int/iris/handle/10665/331921.

WHO, OECD and World Bank. 2018. "Delivering quality health services: a global imperative for universal health coverage." Geneva: World Health Organization, Organisation for Economic Co-operation and Development, and World Bank.

WHO and UNICEF. 2019. "Water, Sanitation and Hygiene in Health Care Facilities: Global Baseline Report 2019." Geneva: WHO and UNICEF.

WHO and World Bank. 2017. "Tracking universal health coverage: 2017 global monitoring report." Geneva: WHO and World Bank.

World Bank. 2020a. World Bank Data Hub. https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD

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.

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

CHAPTER 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.[73]

Before the COVID-19 pandemic, most national education systems already faced a crisis. Its nature was spelled out in the *World Development Report 2018: Learning to Realize Education's Promise* and in the 2013 UNESCO report *The Global Learning Crisis: Why Every Child Deserves a Quality Education*.[74] Despite the expansion of access to schooling in recent decades, the majority of 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 ten-year-olds in the world cannot read and comprehend a simple text.[75]

The pre-pandemic 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.[76] Structural differences associated with poverty, gender, ethnicity, disability, and location explain a substantial portion of schooling disparities.[77] Worldwide, girls were twice as likely as boys to never start school.[78]

The COVID-19 pandemic has exacerbated these pre-existing disparities. At the height of the pandemic, 1.6 billion children worldwide were not physically in school.[79] Coming atop significant family income losses, this unprecedented disruption of education systems has upended learning in many settings and has the potential to scar children's learning and school attainment for years—perhaps

73 United Nations (2011); UN General Assembly (2015).
74 UNESCO (2013).
75 World Bank (2019b).
76 International Commission on Financing Global Education Opportunity (2016).
77 World Bank (2018).
78 Education Commission (2016).
79 United Nations (2020).

generations—to come.[80] Given differences in access to digital devices, internet connectivity, parental involvement, time for supervision, and other factors, the disruption is likely to impact 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 existing 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 shortfall in learning, which are exacerbated by the effects of the COVID-19 shock, innovative solutions are needed to protect learning now while laying foundations for more efficient, equitable, and resilient systems tomorrow. This chapter seeks to contribute to the existing but still incomplete body of evidence in low- and middle-income countries, to inform decision makers and implementers facing tough choices on how to "build back better" in education.

The SDI education sector survey offers a set of indicators for benchmarking the quality of primary education being delivered. In contrast with the health sector, however, the SDI education surveys are also able to directly measure a crucial human capital outcome: student learning for Grade 4 pupils. The assessment of an outcome allows the analysis of SDI in education to show how learning for children is related to the key elements of the provision of education: teachers' knowledge; their effort and time spent teaching; and the availability of necessary infrastructure and materials to teach effectively.

As the SDI health surveys reflect the experience of a typical patient moving through the health care process, so 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 students' learning journey with its progress and pitfalls, 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

80 World Bank (2020c).

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 body 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. 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 these comprehensive surveys is representative at the national level. It is also representative of rural and urban schools and of private and public schools.[81]

As in the health sector, the comparability of the SDI education surveys across countries enables reasonable benchmarking, which in turn can highlight the salience of knowledge and infrastructure gaps as foster momentum for reform. lessons to be drawn across multiple countries in Africa. 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 school children in Sub-Saharan Africa.[82]

The school sample in each country is usually drawn from the national school census or a similar list facilitated by the government. This process usually builds on a dialogue with the 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 over 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 Mozambique, Tanzania, Morocco, and Niger, because private schools are only a small fraction of the education system in those countries. Overall, about 70 percent of the schools in the cross-country sample are public.

The SDI education surveys focus on eight key indicators. The education indicators measure fundamental conditions that shape ordinary students' school experience: rates of teacher absence from the school and the classroom; time spent teaching or time on task; minimum teacher knowledge; minimum infrastructure availability; minimum classroom equipment availability; share of pupils with textbooks; and pupil-teacher ratio. Additionally, student learning is measured through a student assessment on math, language, and non-verbal 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 Grade 4 are tested on math a language, and pedagogical knowledge. Additionally, one Grade 4 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 non-verbal reasoning. Both the teacher

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

82 This sample focuses on SDI surveys that had been fully completed between 2012 and 2018. The 2010 Senegal and Tanzania SDI pilots are excluded because they were conducted at a smaller scale and are not fully comparable. Similar exercises such the Systems Approach for Better Education Results (SABER)–Service Delivery surveys are not considered for this booklet, as 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 based on total country populations from the World Bank DataBank, 2019. Morocco is not included in this estimation as it is not part of Sub-Saharan Africa.

TABLE 2 SDI EDUCATION SAMPLE SIZE BY COUNTRY

Country	Year	Sample		
		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

Notes: This table shows the number of schools, teachers and pupils included in each SDI survey by country. The number of teachers include every teacher in the teacher roster, from which a sub-sample was tested, another sub-sample selected for the absenteeism module, and a smaller proportion observed in the classroom.

and student assessments are based on content derived from the Grades 3 and 4 national curricula of 13 African countries.[83]

During a second, unannounced visit, the school and classroom presence or absence of up to 10 pre-sampled 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. Unstaffed classrooms are also counted directly by the enumerator during the second visit. 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, education SDI surveys measure student learning and generate findings in three key areas:

- **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?
- **Provider's knowledge:** Do teachers have the minimum math, language, and pedagogical knowledge to teach effectively at Grade 4 level?
- **Inputs:** Do schools and classrooms have the necessary equipment (e.g., functioning blackboard and chalk and pens, pencils, and exercise books in fourth-grade classrooms) and infrastructure (e.g., 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 that incorporates private schools in a subset of countries.

83 See Technical Appendix 3 for details on the methodological groundwork for the SDI teacher and student assessments.

Harnessing SDI data to improve systems

The analysis here builds on important previous work using SDI education data which analyzed the state of learning and its associated factors across African countries.[84] Most notably, Bold et al. (2017)[85] studied teacher absence, time spent teaching, teacher knowledge, and pedagogical skills in primary schools using data from eight SDI surveys.[86] The paper documents that on average 44 percent of teachers were absent from class and/or school when they were 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 were absent from the classroom. This paper documents that only two-thirds of teachers know at least 80 percent of the fourth-grade curriculum, with large variation across countries. While over 90 percent of teachers in Kenya and Uganda master the knowledge that their students are supposed to learn, only 25% of Nigerian teachers do. There is an equally concerning level of pedagogical skills. More recently, Bold et al. (2019) used 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, this analysis 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, both in terms of learning and associated factors.

84 Bashir et al. (2018); Bold et al. (2017); Bold et al. (2019); and Mbiti (2016).

85 Bold et al. (2017).

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

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 Grades 1 to 6, although certain schools also offer some pre-primary levels. On average, each school has about 390 children in the primary level, and enrollment is equally distributed 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 is 16:1.[87]

A typical student in Grade 4 in the SDI sample is about 11 years old and equally likely to be a boy or a girl, except in the four surveyed Nigerian states, where student populations showed a male-to-female ratio of 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 will find her teacher absent. When present, teachers are likely to be underprepared. The average teacher in the SDI sample answers correctly to 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, though most classrooms will 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 Grade 4.

Low student achievement overall

Cognitive skills play an important role in explaining educational attainment,[88] and acquiring them early makes it easier to improve them later.[89] 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 4th grade will likely affect many important outcomes in her life. Later sections of this chapter will look in detail at features of the school context that previous literature (e.g., WDR 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 19**). When tested with the SDI student assessment, less than half of children can read a simple sentence out loud correctly, and less than 40 percent can correctly perform single-digit multiplications. While over 90 percent of children can identify three numbers, less than 15 percent can multiply triple digits. These are simple tasks that children are expected to master by the end of Grade 4. When looking at differences by gender, there are almost no difference in language performance, while male students tend to do slightly better in math.[90]

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 correctly identify three letters 22 percent of the time and correctly match words with pictures 34 percent of the time. In this group of students, every other language question, including reading a simple sentence, was answered correctly less than 10 percent of the time. The typical student at the bottom of the distribution was able to read a simple paragraph and correctly answer a basic understanding question less than 8 percent of the time. Low reading proficiency also has implications for math knowledge and a snowballing effect on later learning.[91]

87 Bustamante (2019).

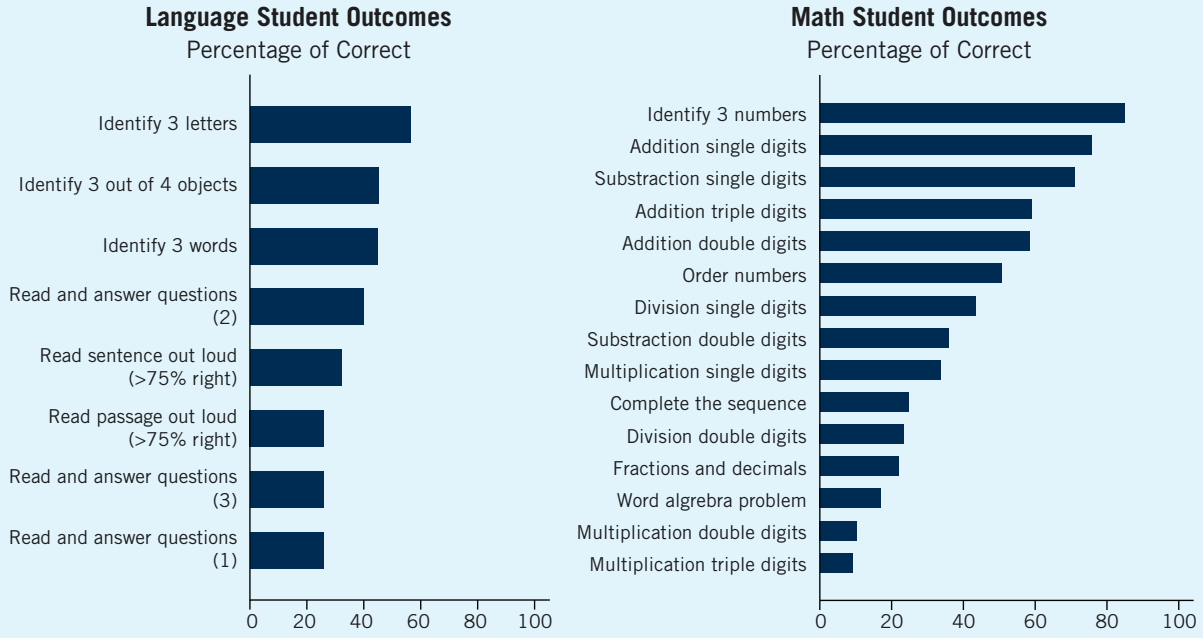
88 Heckman et al. (2006) and Cameron and Heckman (2001).

89 Cunha and Heckman (2007).

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

91 World Bank (2019b).

FIGURE 19 **STUDENT PERFORMANCE IN LANGUAGE AND MATHEMATICS, AVERAGE RESULTS FOR NINE COUNTRIES**



Notes: This figure shows the average percentage of students that correctly answered each task/question on the language and mathematics SDI student assessments. The figures show the simple average across all students in public schools only.

Achievement gaps within and between countries

While the average low levels of learning could appear discouraging, the SDI surveys document substantial student performance differences within and between countries (**Figure 20**).[92] In Niger, for example, the gap between the highest- and lowest-performing students is particularly striking. Indeed, in almost every country there are students performing at the bottom of the overall distribution, as well as students 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,[93] with seemingly three clusters: Madagascar, Morocco, Kenya, and Tanzania at the top; Nigeria, Togo, and Uganda in the middle range, and Mozambique and Niger 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.[94]

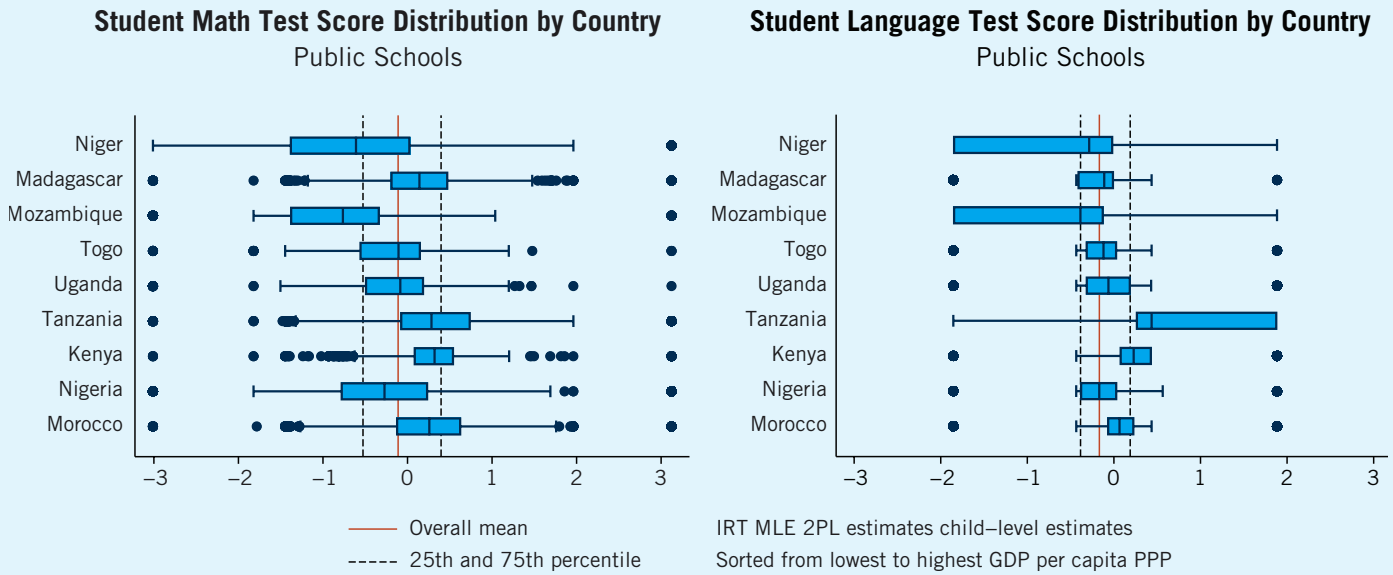
To appreciate the size of this gap, one can think of two groups of students that differ by 1.2 standard deviations in their test scores: a high-scoring group and a low-scoring group. If a student were in the high-scoring group, she would master 8 out of 10 items. If she

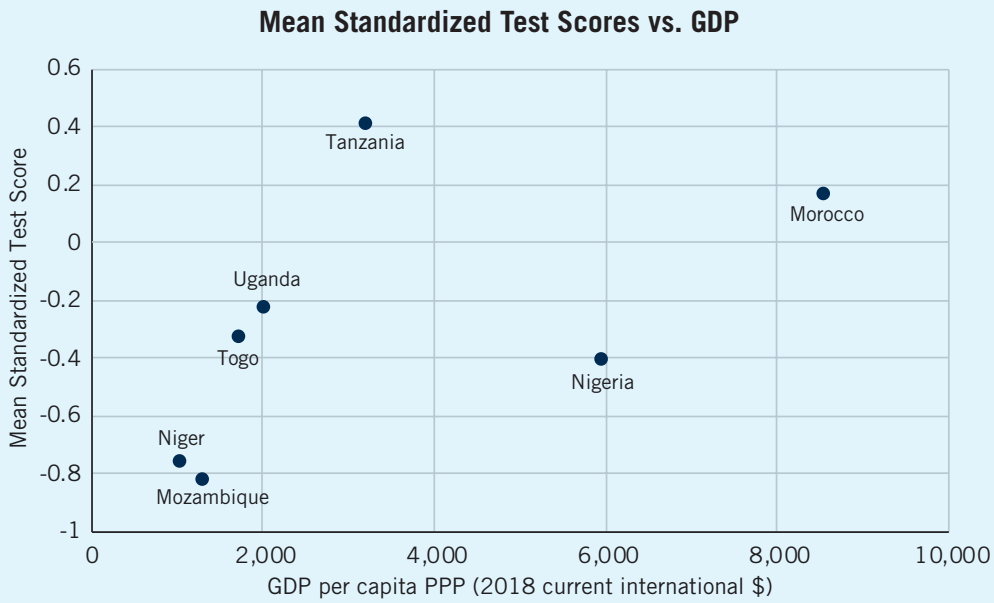
92 For the language component of the student assessment, a sub-sample of students in some countries was tested in a language other than the language of instruction, which was typically the vernacular. While this approach provided important insights for within-country analysis and policy recommendations, 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 5.

93 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 (IRT) and then standardized to have mean = 0 and standard deviation = 1. Maximum likelihood estimates (MLE) are used throughout this booklet. For details on these methods see Das and Zajonc (2010) and Bau, Das, and Yi Chang (2021).

94 This contrasts with Harmonized Learning Outcomes (HLO) rankings and test score differences for this selected group of countries, which positions Kenya with the highest performance and Niger with the lowest performance (and a difference of 1.5 standard deviations).

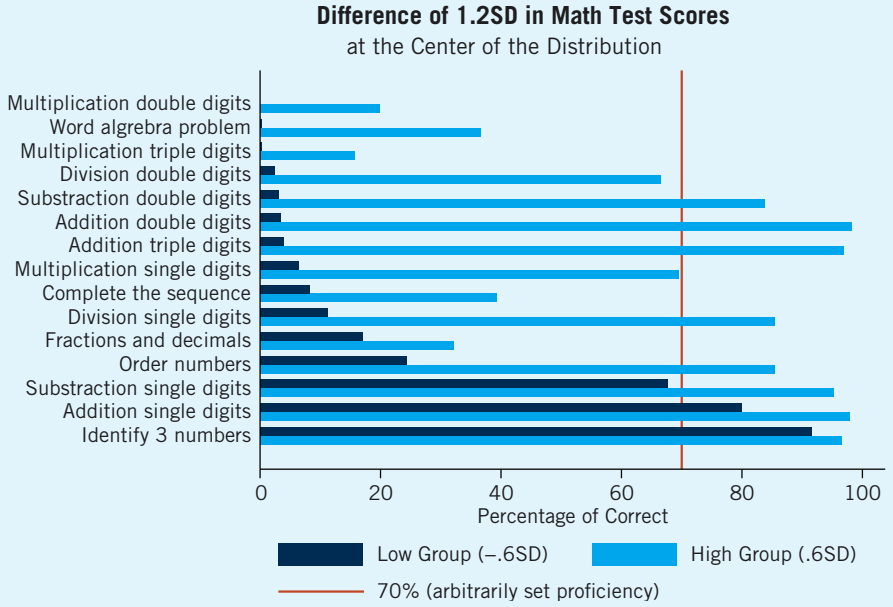
FIGURE 20 STANDARDIZED PERFORMANCE ON THE SDI STUDENT ASSESSMENT, BY COUNTRY





Notes: This figure shows the standardized student test score on the language and mathematics SDI student assessments by country for public schools only. The top panel shows 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 have been equated across countries using an Item Response Theory (IRT) 2 Parameter Logistic (PL) model. The boxplots use Maximum Likelihood Estimation (MLE) results. The solid line represents the mean for all countries and black patterned lines represent the 25th and 75th overall percentiles. Countries are sorted by GDP per capita PPP. The bottom panel shows a scatter plot between the mean standardized student test score combined for both language and mathematics and each countries' GDP per capita PPP.

FIGURE 21 VISUALIZING DIFFERENCES IN TEST SCORES BETWEEN HIGH- AND LOW-PERFORMING STUDENT GROUPS



Notes: This figure shows the mean proportion of correct answers per task/question for the SDI mathematics student assessment for public-school students across all countries combined. Students performing around -0.6 and 0.6 standard deviations are included in the low- and high-performing groups to illustrate the potential difference of 1.2 standard deviations. Note that results might vary if different groups are selected.

were in the low-scoring group, she would only master 2 items. The items that both groups would master are of low difficulty: identifying 3 numbers and addition of single digits. However, the high-scoring group would also master more complex items, such as addition of triple digits and division of single digits, as shown in **Figure 21**.[95]

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 OECD PISA average of 33.6 and 37 percent in 2004 and 2012, respectively,[96] as well as with the 34 percent estimated variation obtained using the World Bank Global Learning Assessment Database (GLAD).[97] While it is possible that students have different experiences in the same school due to factors such as teacher biases, the large amount of unexplained variation suggests that other factors beyond those at the school level might greatly influence learning. Unfortunately, the information available on student characteristics and students' home environment is limited in the SDI sample analyzed here.[98] More recent SDI surveys are attempting to fill this gap (see Chapter 4). The language of instruction may also impact student scores, as detailed in **Box 5**.

Among determinants of student learning, SDI surveys primarily collect information on school inputs and teacher characteristics. For that reason, the remainder of this chapter will focus principally on variation in student learning that can be explained by differences in these characteristics.[99]

95 Another way of understanding this difference is to compare it to the gap in Program for International Student Assessment (PISA) test scores between two countries. A difference of 1.2 standard deviations is equivalent to the gap between Ecuador and the Organisation for Economic Co-operation and Development (OECD) average (OECD 2018) or the difference between Singapore, one of the world's top-ranked countries in math, and Serbia, which ranked 46th (OECD 2019).

96 OECD (2004) and OECD (2013).

97 Azevedo and Goldemberg (2020).

98 A simple ANOVA decomposition shows that, while country-fixed effects explain up to 63.5 percent of the variation, the first principal component obtained through Principal Component Analysis (PCA) of school inputs and teacher characteristics accounts for 6.6 percent and 11.5 percent of the explained variation in the SDI sample.

99 The characteristics that this chapter focuses on were mostly selected 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 (PCA) as explaining the most variation within four category groups of variables corresponding to the WDR 2018 framework.

BOX 5: HOW DOES LANGUAGE OF INSTRUCTION AFFECT TEST SCORES?

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

The literature documenting lower scores is voluminous and consistent. Mullis et al. (2017) found 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.[1] Those not tested in their L1s scored significantly lower, by more than one-third of a SD. Glewwe et al. (2012) found that linguistic minority students fall behind very early in their school experience and have a hard time catching up.[2] Other authors have documented the strong correlation between being taught in one's L1 and continuing in primary school.[3]

Do these persistent 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 among 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 respective L1s, post-test analyses can, in theory, estimate the extent of bias between test versions. In practice, this way of testing and analyzing results is rare outside main languages spoken in OECD countries.[4]

One part of the story is clear: the problem is not the inability of bi- or multi-linguals to achieve as well or better than monolinguals. Collier and Thomas (2017) found that when students receive enough high-quality instruction in both languages, bilinguals will eventually outperform monolinguals even in the monolinguals' L1.[5] 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. ILSAs were initially designed for and first given in OECD member countries, which tend to be more linguistically homogenous than non-OECD countries.[6] 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 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."[7]

The challenge is formidable, but by testing students in their L1s 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 ILSAs. Countries are responsible for translating the assessment into their own languages and adapting it to their respective contexts. In the same spirit, IEA and Boston College conduct studies to detect test and item bias following the standards[8] in the field of psychometrics. 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.

continued on next page

continued from previous page

The growth in participation in ILSAs is an opportunity for many countries and for international development organizations. Organizations behind ILSAs support participating countries with capacity building initiatives, so they can conduct better NLSAs and follow best assessment practices.

Note: Contributed by Michael Crawford.

[1] Mullis et al. (2017).

[2] Glewwe et al. (2012).

[3] Ramachandran (2012).

[4] Ramachandran (2012).

[5] Collier and Thomas (2017).

[6] Ethnologue database. https://www.ethnologue.com

[7] Grisay et al. (2007).

[8] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014) and Educational Testing Service (2014).

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/him throughout life.[100] 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. Teacher absence has been found not only to correlate with lower learning, but causal studies have shown that reducing absence can improve learning.[101] However, even when they are in school, teachers often spend too much time in activities other than teaching. As mentioned earlier, teacher absence in SDI countries is well documented in Bold et al. (2017).[102] 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,[103] 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.[104] Overall, teacher absence remains a substantial challenge for SDI countries included in this booklet. There are many possible reasons behind teacher absence, including system-wide shortfalls in personnel policies,[105] as well as lack of monitoring and accountability and insufficient incentives.[106]

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.[107] Teachers' abilities are often assumed to be associated with academic credentials. However, a growing body of evidence shows that the teacher skills that matter most for learning – content knowledge and pedagogy – are not necessarily linked with teachers' formal qualifications.[108] By providing direct measures of knowledge and pedagogy, SDI surveys make it possible to measure how important teachers' abilities are in explaining children's learning outcomes. In fact, a recent study has also used 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.[109]

100 See discussion in Chapter 1.

101 Duflo et al. (2012).

102 Bold et al. (2017).

103 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 school and classroom absence 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 Tanzania and Senegal in 2010 and from Tanzania 2014, while this booklet excludes those data in favor of newer Tanzania 2016 data and adds data from Niger 2015, Madagascar 2016, and Morocco 2016. This booklet also focuses on public schools (in the current subsection), while the results of the paper 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 booklet can be found in the section on samples, methods, and framework.

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

105 Liu et al. (2020).

106 Muralidharan et al. (2016); Mbiti (2016).

107 See Hanushek and Rivkin (2006), Evans and Popova (2016), and Araujo et al. (2016).

108 Cruz-Aguayo et al. (2017); Rivkin, Hanushek, and Kain (2005); and Hanushek and Rivkin (2012).

109 Filmer et al. (2020).

FIGURE 22 EXTRACTS FROM TEACHERS' ASSESSMENTS

		(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)

		(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) $72 \div 9 =$ 7	(d)	(d)
	(e) $37 \times 13 =$ 371	(e)	(e)
2	Which two numbers add up to make 0.81? 0.24 0.09 0.9 0.51 0.57 0.17		
3	Circle the one that gives the smallest answer? (a) $39 \div 5$ (b) $39 \div 4$ (c) $39 \div 3$		

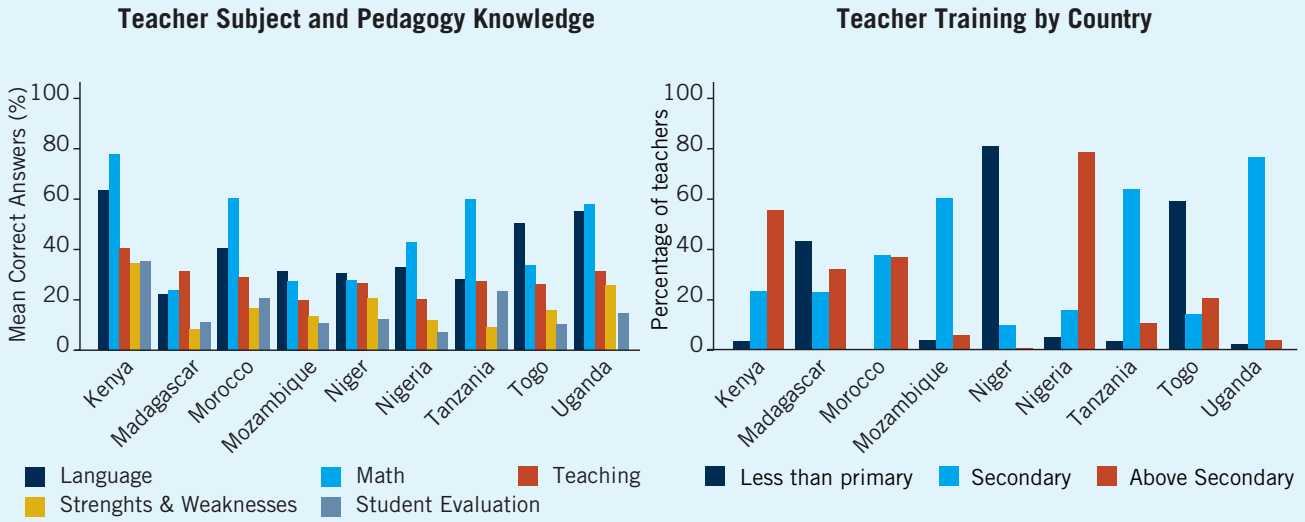
The SDI teacher assessment includes two sections. The teacher knowledge section resembles grading a math and literacy exam (e.g., correct a letter with grammatical errors, grade arithmetic exercises solved by students, and similar tasks), while the pedagogical section ask teachers to perform tasks that they face on a daily basis (e.g., 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 22** are examples of the types of questions on which teachers are assessed.

Consistent with findings from previous work, teacher knowledge and pedagogical skills are also low in this expanded SDI sample. The average teacher in any of the countries cannot correctly answer more than 80 percent and 65 percent of the math and language knowledge assessment questions, respectively. 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 23** shows the average percentage of correct answers by task and the percent of teachers by formal training for each country. Teacher knowledge and skills do not always seem to be correlated with the highest level of formal schooling that teachers have completed (i.e., primary or less, secondary, or above secondary), which aligns with the literature on this subject.[110] For instance, while Kenya has a high average teacher knowledge test score and a high percentage of teachers with an education above secondary, Nigeria has a relatively low teacher knowledge score even with a higher percentage of teachers with training beyond secondary school.

This evidence can be a force for change. An SDI survey in Niger in 2015 found that teacher competencies were weak in comparison with other African countries, and that a small fraction of the teachers tested had the minimum level of knowledge to teach French and mathematics at the primary education level. Local researchers led 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 additional data generation on the quality of teachers, and the Ministry of Primary Education launched a large-scale national assessment of primary school contractual teachers (who represent 75 percent of the

110 A regression at the teacher-level of teacher assessment scores and teacher education level groups finds significant and positive albeit very small associations (2-4 percentage point differences) between level of education groups and teachers' knowledge and pedagogic skills after controlling for country fixed effects, urban/rural locality, teachers' age and gender.

FIGURE 23 TEACHER SUBJECT AND PEDAGOGY KNOWLEDGE AND TRAINING



Notes: The figure on the left shows the percentage of correct answers per task/question for the SDI teacher assessment for public-school teachers, sorted by country. The figure on the right shows the percentage of teachers by country with different levels of training. Categories do not always add to 1, as there are some missing.

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 impact learning when they support productive teacher-learner interactions.[111] 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 features are not always available to students in low-resource settings. Factors like student-teacher ratios are also key. Learning will not be as effective when classes are too large or extremely diverse in terms of school readiness.[112] 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.[113] These infrastructure characteristics also represent the working conditions for teachers – it is extremely difficult for teachers to apply the skills they have when these basic inputs are not there.

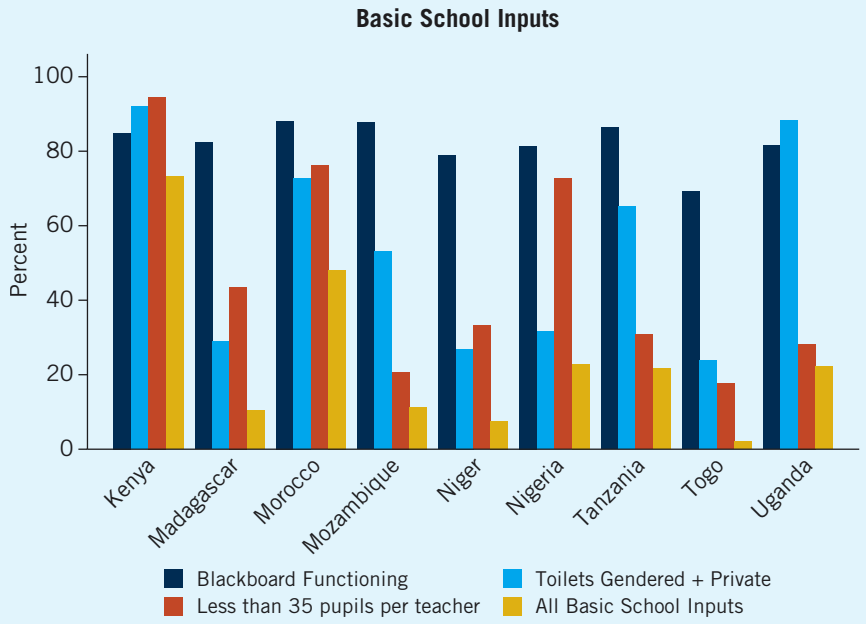
The SDI data allows visualization of the average percentage of schools by country that have some basic school inputs (**Figure 24**). While most schools across all SDI countries have a functioning blackboard (i.e., a board with chalk that is visible to all students), the number of schools with toilets 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 around 20 percent in 6 out of the 9 countries in this SDI sample. This low percentage shows that there is considerable space for government action to equip every school with basic inputs that might help improve the quality of education.

111 World Bank (2018).

112 Ganimian and Murname (2016).

113 Adukia (2017).

FIGURE 24 BASIC SCHOOL INPUTS



Notes: The figure shows the percentage of public schools by country that had each basic school input and all combined.

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. While these are not the only factors that influence learning, there is evidence that teacher quality and certain school inputs matter in specific contexts.[114] On the other hand, especially in low- and middle-income African countries, evidence is currently 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—was found to be markedly low in a subset of SDI countries.[115] It is plausible that prioritizing improvements to 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 with additional school inputs.[116] Ultimately, identifying areas of potential action to improve student experience and learning can help strengthen education systems in low- and middle-income countries in Africa and elsewhere. This is an important direction for future analysis.

114 World Bank (2018); Hanushek and Rivkin (2006); Evans and Popova (2016); and Araujo et al. (2016).

115 See Bold et al. (2017), which also uses SDI data.

116 Mbiti et al. (2019).

LEARNING FROM WHAT WORKS: WHAT CAN COUNTRIES LEARN FROM THEIR OWN SUCCESS

Improving educational systems in a relatively short period of time is challenging but feasible and there are many success stories.[117] 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 boost learning outcomes.

Homing in on 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 yields a group of schools where students are performing well and another group where students are performing poorly. In the best-performing schools (top 5 percent) the average student masters at least 9 out of 15 math items, while in the lowest-performing group (bottom 5 percent) the average student masters none.[118]

The best-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 best-performing schools are clustered in five countries, led by Kenya, Tanzania, Morocco, and Madagascar. Interestingly, Nigeria has a sizeable share of both high-performing and low-performing schools, indicating a wide variation in school quality within the country.

Studying the characteristics of these high- and low-performing schools in SDI countries, some noteworthy patterns arise. 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 standard deviations higher test scores for urban schools).[119] 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.[120] The best-performing schools do significantly better on almost every school input. For instance, schools in the best-performing group are 20 percentage points more likely to have a functioning blackboard with equipment to write on it, visible from all seats within the classroom. In terms of pupil-teacher ratio, the best-performing schools have around 35 students per teacher, while their worst-performing counterparts have an average of 56 students per teacher (see Appendix Table A5 for details).

Best-performing schools not only have more teachers per student than their counterparts in the worst-performing group, but their teachers are also better prepared. Teachers in the best-performing schools scored 1.3-1.4 standard deviations higher on numeracy and literacy knowledge, respectively, and 0.65 standard deviations higher on pedagogy knowledge. The top group of schools also retain teachers who are on average older by about 4 years and more likely to be female by 40 percentage points.

117 A substantial number of such stories are documented in World Development Report 2018 (World Bank 2018).

118 "Mastering" refers to having more than 70 percent of correct answers across all students in the school groups.

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

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

What may be driving differences in school performance?

Comparing high- and low-performing schools, Appendix Table A5 highlights the specific teacher characteristics and school inputs that appear to most clearly distinguish the best-performing schools from those 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 have led one to predict.[121] But other, 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 emerged 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 best- and worst-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. This 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 that may help target interventions to improve school inputs. For instance, it is possible to look at how many schools per country and/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, while 35 percent of urban schools meet this basic standard. Thus, in some settings, targeting rural schools early for additional inputs could make sense. In some countries, on the other hand, SDI data may serve as a more generalized wake-up call to decision makers on school inputs. 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 on school inputs. Yet, caution is warranted, since 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.[122] On average, other factors remaining constant, adding one teacher per school and ensuring that every school has a functioning board and private and gendered toilets could be associated with an increase of 0.24 standard deviations in students' average test scores (see Appendix Table A6).[123] While this evidence does not give insights into causal pathways for reform, it still points at potential and feasible solutions with clear benefits in the order of magnitude of the impact generated by the most successful interventions to improve test scores at scale.[124] Some of these improvements, such as the presence of handwashing facilities in schools, acquire renewed importance in the wake of COVID-19 (see **Box 6**).

121 Chetty et al. (2011). Note that this does not necessarily contradict the results of Duflo, Dupas, and Kremer (2015). The latter researchers found that lowering class size by adding more centrally hired civil service teachers did not improve student learning outcomes, but they attributed this null effect to existing teachers' reducing their effort in response to the new hires and helping their relatives get hired into a significant portion of the new teaching slots.

122 Ganimian and Murnane (2016).

123 Result from a multivariable linear regression of student test scores on school inputs controlling for country-fixed effects and proxies for accessibility to school and poverty.

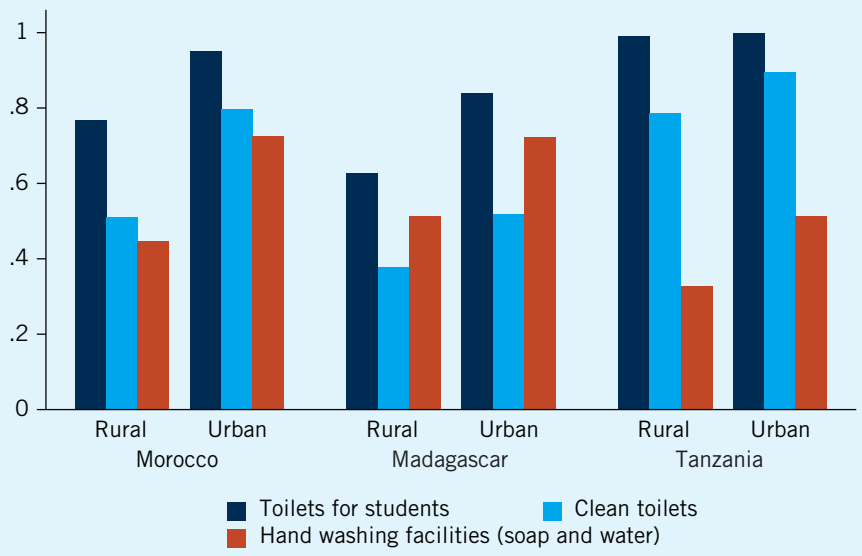
124 See Glewwe and Muralidharan (2016); McEwan (2015).

BOX 6: TOILETS AND HANDWASHING FACILITIES IN SCHOOLS: A KEY CONCERN DURING COVID-19

In many respects, the COVID-19 pandemic has reshaped the way people think about education. Millions of schools have shut down and, while some later reopened, many are still assessing whether it is safe to do so. An essential condition to allow schools to reopen is ensuring a safe return for students and teachers. This 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 will also be essential to prevent and/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 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. While some SDIs date back several years, more recent ones – including Tanzania, Morocco, and Madagascar surveys from 2016 – may still give policymakers a good indication on what investments are needed in this area to facilitate a safe reopening. In particular, the data allows analysts to look into: (i) the availability of toilets for students; (ii) the availability of a clean toilet; and (iii) the availability of a handwashing facility with soap and water near the toilets.

FIGURE 25 TOILETS AND HAND WASHING FACILITIES



Notes: This figure shows the percentage of facilities with different pieces of personal protective equipment (PPE) available in Sierra Leone, with the dots showing discrepancies by facility type.

continued on next page

continued from previous page

SDI data shows that, generally, sanitation infrastructure investments are needed more urgently in rural settings, although schools located in urban areas will also need support. While 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 26** shows. However, there is more space for improvement when the bathroom as directly observed by the enumerator is required to be clean. Across the three SDI 2016 countries, only 65 percent of schools appear to have a toilet assessed as clean.

Of particular importance in the setting of COVID-19 is the presence of handwashing stations near these toilets. As shown in the figure above, this feature may not have been prioritized in the past. Overall, only 48 percent of schools in the three countries seem to have a handwashing station in proximity to their sanitary facilities. That means that more than 50 percent of students will not be able to wash their hands after using the toilets. This 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 make it possible to shed some light on this and other characteristics that can help policymakers 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 5 explores some future survey plans in greater detail.

Leveraging teacher traits to improve results

The comparison exercise between the best-performing and worst-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 on pedagogical tasks they should perform on a regular basis.[125] Other characteristics that are easier to observe, such as formal education and accreditation, have been shown to have no link to better teacher performance.[126]

There is substantial variation in teacher characteristics across SDI countries. On average across all countries, 13.6 percent of teachers have an education of primary or less, 50.5 percent have completed secondary schooling, and 35.4 percent have completed a post-secondary degree.[127] Niger has the highest percentage of teacher with a primary education or less (86.2 percent), while Nigeria has the most educated teachers, with the vast majority having a post-secondary degree (80.3 percent). Large differences in pedagogy, numeracy, and literacy between the best- and lowest-performing schools can be observed. Surprisingly, teacher effort does not seem to be a critical factor in learning outcomes, though this may be because the proxy measure for effort – teacher absence – only captures certain dimensions of effort. On the other hand, teachers' knowledge is a strong predictor of student learning. Consistent with the literature, teacher quality does not seem to be strongly linked to teachers' formal education and accreditations in the SDI data.[128]

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

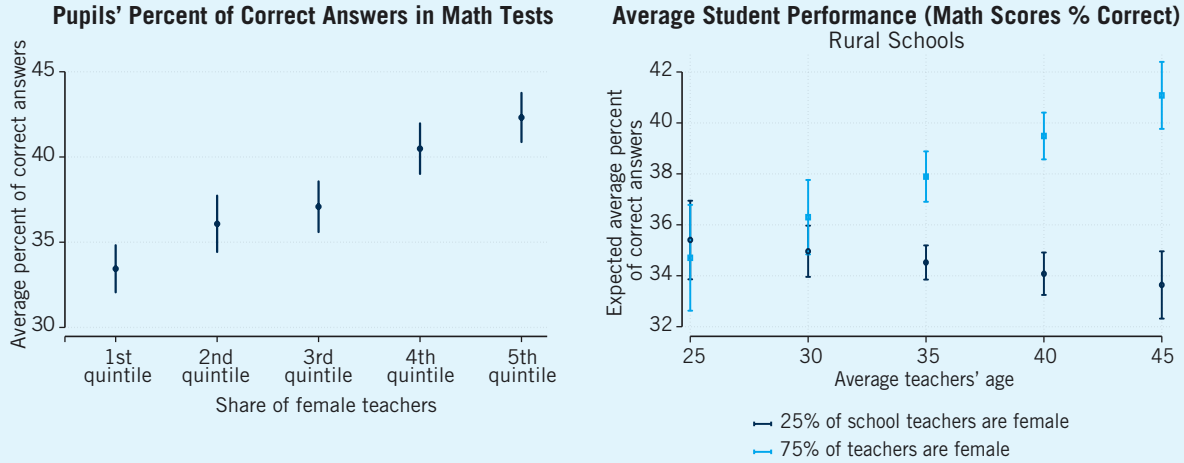
125 See, for instance, Hill et al. (2005); and Metzler and Woessmann (2012).

126 Bau and Das (2020); Hanushek and Rivkin (2006).

127 The remaining 0.5 percent corresponds to the category "Other."

128 A regression at the teacher-level of teacher assessment scores and teacher education level groups finds significant and positive albeit very small associations (2-4 percentage point differences) between level of education groups and teachers' knowledge and pedagogic skills after controlling for country fixed effects, urban/rural locality, teachers' age and gender.

FIGURE 26 TEACHER GENDER AND AGE ARE ASSOCIATED WITH LEARNING PERFORMANCE IN SDI COUNTRIES



Notes: The figure on the left shows the average percentage of correct answers for the SDI math student assessment for students in public schools, grouped by within-country quintiles according to the share of female teachers in schools. The figure on the right shows the linear prediction for the average percentage of correct answers for the SDI math student assessment for students in rural public schools, grouped by the average teachers' age in the school for schools with 25 percent and 75 percent of female teachers. The estimates come from a regression controlling for teachers' formal training and country fixed effects. Each linear prediction shows 95 percent confidence intervals with robust standard errors.

perform better across the SDI sample of public schools. This positive association is driven in great part by schools located in rural areas and where female teachers are older. One potential explanation behind this association is that schools with higher shares 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 above in **Figure 26**.

Across the SDI sample, older female teachers perform better than their counterparts on key measures that correlate positively with pupils' test scores, including teacher pedagogy, numeracy, and literacy scores as well as lower teacher absence rates. These correlations are stronger in rural settings.[129] In some countries, teaching for a long time represented one of the few career paths deemed culturally appropriate for women to access the labor market.[130] This might explain why a relatively large number of qualified, motivated women entered the teaching profession, and this is reflected in generational trends. While 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 the other, even within the SDI sample. In Uganda, for example, a share of private schools receives some government funding and, as a result, are subject to some rules and regulations, while 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

129 These results come from regressions at the teacher-level controlling for school locality, teacher formal training, and country fixed effects.
130 Goldin (2015).

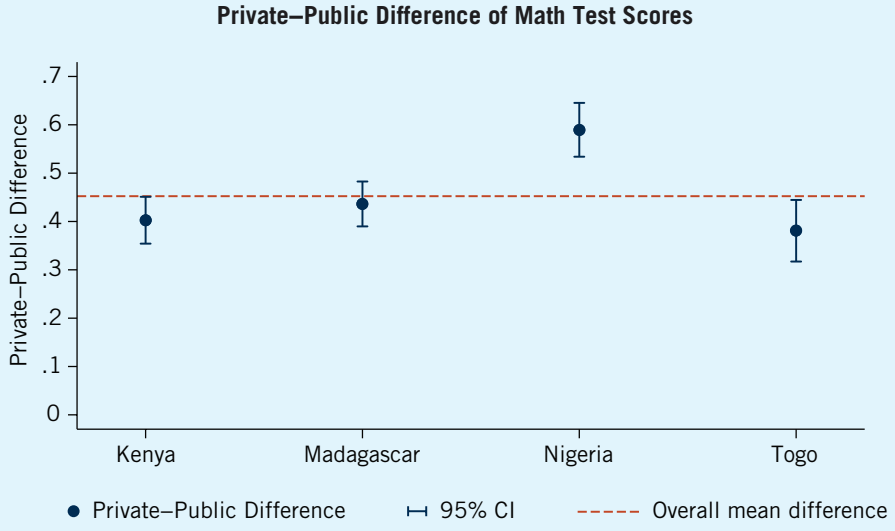
schools were not included in the sample frame due to their scarcity. For the same reason, in Morocco, there were virtually no rural private schools in the final sample. Thus, comparing private schools in either of these two countries with those in other SDI countries would implicitly limit one side to only urban schools.

Notwithstanding these challenges, studying the private educational 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 sub-sample of countries where both private- and public-school sectors are sufficiently large 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 represented around 60 percent to 78 percent of each country sample, and the remainder schools were classified as private. The composition of the private sector, however, varied depending on the context. For instance, in Kenya, 15 percent of schools were private for-profit and 7 percent private not-for-profit, with the remaining 78 percent being public. In Madagascar, 64 percent of schools were public, 12 percent private with religious affiliation, and 23 percent private without religious affiliation. In Nigeria, 3 percent of schools in the sample were private community schools, 7 percent private not-for-profit, and 30 percent private for-profit, with the remaining 60 percent being public. Finally, in Togo, 13 percent of the sample included private schools without religious affiliation, 14 percent were private with religious affiliation, and the remaining 73 percent were public schools. Despite these differences, some general trends emerge when comparing the private and public sectors.

Results, shown in **Figure 27**, indicate that, in Kenya, Madagascar, Nigeria, and Togo, private schools do better than public schools, on average, as captured by better performance in the math test by children attending private schools. The exercise previously described of grouping best- and worst-performing schools was repeated for this subset of countries, but now including private schools. Some suggestive patterns emerge from this analysis. First of all, there are different types of high-performing schools among all the possible combinations of urban/rural and private/public groups of schools. Urban private schools make up the largest single share of the best-performing schools (40 percent). There are almost no urban private schools and very few rural private schools among the worst-performing schools.

FIGURE 27 DIFFERENCES IN MATH TEST SCORES BETWEEN PUBLIC AND PRIVATE SCHOOLS, FOUR COUNTRIES



Notes: This figure shows the average difference between SDI math student assessment results between 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 red dotted line represents the average across this selected group of countries. Ninety-five percent confidence interval shown with robust standard errors.

TABLE 3 COMPARISON OF BEST- AND WORST-PERFORMING SCHOOLS IN SELECTED COUNTRIES AND INCLUDING PRIVATE SCHOOLS

	Bottom				Top			
	Rural		Urban		Rural		Urban	
Variable	Public	Private	Public	Private	Public	Private	Public	Private
Test Scores	-2.10	-1.59	-1.59	-1.11	1.86	1.62	1.43	1.56
Teacher Characteristics Principal Component Analysis (PCA)	-0.54	-0.72	0.15	-1.31	-0.12	0.50	-0.95	0.00
School Inputs PCA	-1.25	-0.89	-1.32		0.05	0.49	-0.09	0.66
N	141	19	13	1	48	48	8	69

Notes: This table shows the number of schools, teachers and pupils included in each SDI survey by country. The number of teachers include every teacher in the teacher roster, from which a sub-sample was tested, another sub-sample selected for the absenteeism module, and a smaller proportion observed in the classroom.

Within the lowest-performing group of schools, the few private schools present are doing better than public schools (**Table 3**). However, among the best-performing group, public schools are doing as well or better than private schools. While 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.

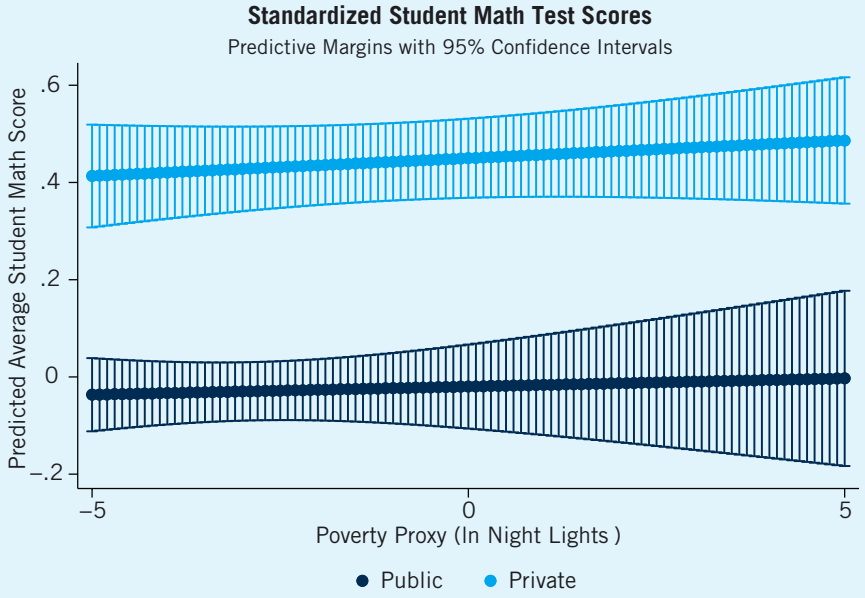
One possible reason why the private schools do overall better in terms of learning is that they might have a higher share of wealthier students, who can potentially take advantage of better conditions for learning at home. In this case, the difference in learning outcomes between public and private schools 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. While the most recent SDI surveys are now collecting this information, previous SDI surveys have geo-references which may be merged with geospatial data to start addressing some of these questions (see Chapter 4 for details). In this case, night lights can be used as a proxy for the socioeconomic status of the school location.

Expectedly, students in more affluent locations do better, but the difference between private and public schools stay constant throughout the whole welfare spectrum (**Figure 28**). This is suggestive that the difference between public and private schools observed in the sample cannot be entirely explained by the socioeconomic composition of their student body. Nonetheless, the same patterns could be observed in the presence of within-area sorting of students and more accurate measures of students' household socioeconomic status would help us to gain clearer insights in the future.

There are many potential drivers behind the difference 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.[131] Better management could potentially contribute to explaining the public-private gap in learning outcomes observed in this sample. Unfortunately, as with household socioeconomic status, this information has historically not been collected by SDI surveys. Chapter 5 will describe some recent SDI efforts to better capture these two dimensions. The new approaches hold promise to clarify this and other policy-relevant correlations with a view to improving student learning outcomes in low- and middle-income contexts.

131 OECD (2012).

FIGURE 28 **PREDICTION OF STUDENT OUTCOMES OVER THE RANGE OF A PROXY FOR SOCIOECONOMIC STATUS BY PUBLIC AND PRIVATE SCHOOL OWNERSHIP**



Notes: The figure shows the linear prediction for the SDI math student test scores across the whole distribution of night lights (a proxy for poverty), controlling for urban/rural locality, a wide array of teacher characteristics (e.g. age, sex, teacher formal training, etc.) and country fixed effects. Each linear prediction shows 95 percent confidence intervals with robust standard errors.

CONCLUSIONS: WHAT WILL IT TAKE TO IMPROVE SERVICE DELIVERY IN EDUCATION?

A comprehensive review of all the policies that are necessary to improve the quality of education, during and after COVID-19, is beyond the scope of this report. Various WBG flagship reports have surveyed the entire 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.[132] That said, SDI evidence offers unique insights into margins that could significantly help improve students' experience and their learning, particularly in the African context. In combination with previous evidence in the field, this analysis can be translated into practical guidance to 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 the following broad directions for action. Each country can and should shape the principles to its own specific context and priorities. A key step is recognizing what is already working in each national setting.

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 due to cultural and institutional differences. Extrapolating lessons from top performing public schools might represent an easier first step for countries that want to improve their learning outcomes. By providing comprehensive information on education systems, SDI surveys can help governments deploy investments so that larger numbers of young people can reach their full potential.

132 World Bank (2018, 2019b, 2020c).

To improve inputs, start with the basics—but don't stop there. Consistent with findings from previous impact evaluations,[133] SDI results point to the importance of specific school inputs that can potentially improve students' daily experience in school, such as whiteboards and functioning toilets. These investments, while 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 could be a good place to start. On the other hand, in most cases, such inputs alone will probably 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 appear to have more teachers per student, and also more knowledgeable teachers. Accordingly, policymakers looking to make the most effective and efficient use of resources may also be looking to hire more and better teachers. Among a mixed literature that suggests that, in low- and middle-income countries, lowering the pupil-teacher ratio might be effective only under certain circumstances[134] and, if incentives are aligned and coupled with the necessary complementary interventions,[135] class size emerges as a strong correlate of performance differences between schools in the SDI sample.

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

Support efforts to encourage teachers to be learners. Perhaps one of the most challenging tasks for education authorities will be implementing 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 seem to display higher average levels of learning than public schools. As a result, much can potentially 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 alternative solutions. For example, while private schools might be able to attract better school principals through 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 then contributing to better learning outcomes.[136]

The COVID-19 pandemic and its ensuing economic crisis will likely 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 is available, countries can and have been taking steps to identify and target interventions directed at students at high risk of dropping out, or who are lagging behind on learning. For instance, teaching at the right level or remedial tutoring will be 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 bridge this gap and support policymakers 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 will require strategic leadership informed by evidence and analysis of the main drivers of learning. Consolidating resilient educational 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 leapfrog advances for students, teachers, families, and societies. The next chapter looks in detail at some of the measurement innovations that can drive this work.

133 See Ganimian and Murnane (2016) for a literature review.

134 Chetty et al. (2011); Duflo et al. (2012); and Ganimian and Murnane (2016).

135 For instance, Mbiti et al (2019) shows in Tanzania significantly larger effects on student test scores of a joint intervention of teacher incentives coupled with school grants versus incentives or grants provided separately.

136 Muñoz and Prem (2020).

REFERENCES

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

American Educational Research Association, American Psychological Association, 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): 1415–1453.

Azevedo, J. P., and D. Goldemberg. 2020. "Learning for All: Within-country learning inequality." World Bank Blogs - Education for Global Development. 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 [Forthcoming]. "Learning Trajectories through Primary Schools: New Evidence from Pakistan." *International Journal of Educational Development*.

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." World Bank Policy Research Working Paper 8849, World Bank, Washington, DC.

Bustamante, J. 2019. "K-12 School Enrollment & Student Population Statistics." Educationdata.org. https://educationdata.org/k12-enrollment-statistics#

Cameron, S.V. and Heckman, J.J., 2001. "The dynamics of educational attainment for black, hispanic, and white males." *Journal of Political Economy*, 109(3), pp.455-499.

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." *The Quarterly Journal of Economics* 126(4): 1593-1660.

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: 203-217.

Cruz-Aguayo, Y., P. Ibarrarán, and N. Schady. 2017. "Do tests applied to teachers predict their effectiveness?" *Economics Letters* 159.

Cunha, F. and Heckman, J., 2007. "The technology of skill formation." *American Economic Review*, 97(2), pp.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., Dupas, P., & Kremer, M. (2015). "School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools." Journal of Public Economics, 123, 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.

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

Educational Testing Service. 2014. "2014 ETS standards for quality and fairness." Princeton, NJ: Educational Testing Service.

Ethnologue database. 22nd Edition. https://www.ethnologue.com

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

Filmer, D., E. Molina, and W. Wane. 2020. "Identifying Effective Teachers: Lessons from Four Classroom Observation Tools." The 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): 719-755.

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." Young Lives Working Paper No. 80., 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 Eric A. Hanushek, Stephen J. Machin, and Ludger 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,* 313-348. 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): 249-266.

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

Hanushek, E.A. and Rivkin, S.G., 2012. The distribution of teacher quality and implications for policy. Annu. Rev. Econ., 4(1), pp.131-157.

Heckman, J.J., Stixrud, J. and Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. Journal of Labor economics, 24(3), pp.411-482.

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." New York: International Commission on Financing Global Education Opportunity.

Liu, J., S. Loeb, and Y. Shi. 2020. "More Than Shortages: The Unequal Distribution of Substitute Teaching." *Education Finance and Policy*: 1-48.

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, R. Rajani. 2019. "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania." *The Quarterly Journal of Economics*, Volume 134, Issue 3, August 2019, Pages 1627–1673, 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): 353–394.

Metzler, J. and Woessmann, L., 2012. The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. Journal of development economics, 99(2), pp.486-496.

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

Muñoz, Pablo, & Mounu Prem. 2020. "Managers' Productivity and Labor Market: Evidence from School Principals." Working paper 40, Red Investigadores de Economía.

Muralidharan, K., Das, J., Holla, A. and Mohpal, A., 2016. The fiscal cost of weak governance: Evidence from teacher absence in India. The World Bank.

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. 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. 2013. "PISA 2012 Results: What Makes Schools Successful (Volume IV): Resources, Policies and Practices." Paris: OECD. https://doi.org/10.1787/19963777

OECD. 2018. "PISA for Development: Results in Focus." *PISA in Focus* 91 (December).

OECD. 2019a. "PISA 2018 Results (Volume I): What Students Know and Can Do." Paris: OECD.

OECD. 2019b. "PISA 2018 Results (Volume II): Where All Students Can Succeed." Paris: OECD.

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

Rivkin, S.G., E.A. Hanushek, and J.F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73: 417-458. https://doi.org/10.1111/j.1468-0262.2005.00584.x

UN General Assembly. 2015. "Transforming our World: the 2030 Agenda for Sustainable Development." 21 October 2015, A/RES/70/1. 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.

United Nations. 2011. "Millennium Development Goals Report 2011," June 2011, ISBN 978-92-1-101244-6, available at: https://www.refworld.org/docid/4e42118b2.html

United Nations. 2020. "Policy Brief: Education during COVID-19 and beyond." August 2020. https://unsdg.un.org/resources/policy-brief-education-during-covid-19-and-beyond

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

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

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

CHAPTER 4

AN EVOLVING MEASUREMENT AGENDA FOR THE QUALITY OF SERVICE DELIVERY

For the past decade, SDI surveys have provided evidence to policy makers, implementers, and communities seeking to improve the quality of health and education services in low- and middle-income countries. Now, COVID-19 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 report 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 front-line health and education service delivery. In doing so, it reflects changes in thinking about service delivery and in measurement technology that have emerged over the past 10 years. The chapter first discusses adaptations that are enabling new SDI surveys to better reflect specific country needs and policy priorities, while preserving the comparability of core indicators. It then describes new measurement approaches that better capture: (1) 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 is 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 better speak to country contexts; new health-facility survey modules that measure patient satisfaction; questionnaires that capture data on job satisfaction, mental wellbeing, and soft skills among health workers and teachers; and new modules that help 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 child growth delay, reflecting cumulative insults to a child's health and nutrition. Stunting continues to be a global problem and, in countries such as Madagascar and Guatemala, stunting rates approach 50 percent.[137] A prior SDI survey in Madagascar offered some recommendations on the stunting issue, along with recommendations for improving the education system. Stunting has many determinants, including household ("demand-side") risk factors such as lack of safe water, sanitation, and hygiene; use of biomass fuels; or non-exclusive breastfeeding practices.[138] Just as important, but less discussed, is the role of health care providers (the "supply side") in detecting growth delays, advising about healthful practices, and delivering high-impact interventions to prevent and manage malnutrition. While population-based risk factors and their associated burdens are commonly measured in household surveys, no facility survey exists at present to comprehensively measure supply-side capacity to detect malnutrition among patients and 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 growth delays and advise caregivers about appropriate nutrition practices. Health care providers are tested on their ability to detect stunting during a simulated routine child check-up. 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 Integrated Management of Childhood Illness (IMCI) guidelines to assess the child's symptoms, measure their growth, and decide on appropriate next steps. The health care provider should be able to pick out 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 data on the household side. This is the second key dimension of some of the new-generation SDI surveys. To collect these data, households in Guatemala were sampled within the catchment area of a health facility, with oversampling of households containing women of childbearing age who are more likely to have young children. Household questions focus on access (and barriers) to health services, caregivers' and mothers' behavior in terms of 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 (CHWs), the SDI survey in Madagascar is being adapted to include a separate module to measure CHW knowledge about antenatal care and maternal and child nutrition.

137 USAID Madagascar nutrition profile (2018); USAID Guatemala nutrition profile (2014).

138 Danaei et al. (2016).

As nutrition is important both from 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 included may cover topics such as:

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

Mental health and non-communicable 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 disease burden trends have evolved globally, and the geographic scope of the surveys expands, the SDI surveys need to be modified to fit a broader range of potential health conditions, including non-communicable 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.[139] This process has begun in Moldova, Ukraine, and Bhutan, where new SDI vignettes have been tailored to include cases of depression and/or hypertension.

The WHO's Mental Health Gap Action Programme (mhGAP) intervention guide provides recommendations for treating a variety of mental health disorders in non-specialized health care settings.[140] 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 (DALYs) 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 better integration of NCD prevention programs into primary health care.[141] This 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 provider 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 can lead to serious health complications, such as heart disease and stroke, but is often underdiagnosed in low- and middle-income settings.[142] The health care provider is graded on her ability to arrive at the correct diagnosis, to prescribe appropriate anti-hypertensives, 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.

139 Min et al. (2017).
140 WHO (2019).
141 Kruk et al. (2015).
142 Kim and Radoias (2018).

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 in turn affect the demand for and utilization of education and health services. Nevertheless, correlational evidence between observed or perceived poor quality of services and bypassing or underutilization has been noted in both the education and health literature. For instance, research in Kenya finds that the quality of maternal and newborn care available to the poor is significantly worse than that for the better off,[143] whereas in Tanzania, women bypass their nearest primary care facilities to deliver at more distant health facilities, citing concern about the quality of care at government dispensaries and health centers.[144] Likewise, education studies in the United States[145] and Israel[146] have shown a strong negative correlation between class size in early primary grades and, respectively, college attendance and student learning outcomes, suggesting that low quality schooling has lingering effects on student performance and future enrollment.

Several ongoing SDI initiatives will help address this methodological and evidence gap in both the education and health fields. First, research efforts are currently underway to merge data collected through the SDI education and health surveys in Kenya with existing geospatial data on cell phone connectivity, road infrastructure, nighttime lights, and population characteristics. Data linking will open new avenues for understanding 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.[147] Individuals living within the catchment area of a health facility 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 school children/parents) on perceived quality, current service utilization patterns, and intended future use. These data, joined to those obtained on the supply side, will better enable future exploration of linkages between the provision of services and their utilization.

The new generation of SDI surveys do more to capture people's experiences with health and education services. For the health survey, this 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 him/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 potential side effects of medications, and promoted risk-factor mitigation 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 patient assessment of satisfaction with care, these new questionnaire features anchoring vignettes that allow quantification of respondents' expectations and thus adjustment for subjectivity.[148] 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 there is a large share of variation in student learning that cannot be explained by school characteristics. In order to better understand factors associated with this unexplained variation, novel household and child questions are being explored with the objective of better capturing students' and families' experience with education systems and determinants of education investments. Among other changes, survey designers are considering new sections on parents' and children's

143 Sharma et al. (2017).
144 Kruk et al. (2009).
145 Chetty et al. (2011).
146 Angrist and Lavy (1999).
147 This survey is currently on hold due to the COVID-19 pandemic but will resume as soon as conditions permit.
148 Larson et al. (2017).

perception of teacher quality and skills, including emotional support skills. For instance, parents are being asked what they think makes teachers good at increasing 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 returns to education can shape parental investment.[149] Questions on parental expectations, as well as parents' 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 a battery of questions about whether the parent thinks that the 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 will are correlated with student learning and school choice is an important new area of research.

The World Bank has recently launched the Global Education Policy Dashboard, which integrates key SDI indicators into a system-wide 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 Education Policy Dashboard has developed a cost-effective way to assess student vocabulary in first grade, that the SDI team will build upon 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 the Satisfaction of Employees in Health Care (SEHC) survey, which has been previously validated in low- and middle-income settings.[150] The workload section is designed to measure the burden of both clinical care and administrative tasks, responding to concerns around 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 give an indication of the work environment and to highlight areas where it could be improved.

In education, SDI surveys have adapted to collect data on mental well-being (particularly in fragile and conflict contexts) and on socio-emotional 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 education SDI survey in Indonesia (2019) included a teacher module on socio-emotional skills and a household module for parents aimed at assessing parental school choice and socio-emotional conditions. Both the teacher and parent modules captured self-reported measures of socio-emotional skills and included sections with items derived from the Item Grit Scale, The Big Five Inventory, and the Growth Mindset.[151] Preliminary results will be shared in the forthcoming Indonesia SDI report 2021.[152]

149 Bergman (forthcoming). Attanasio and Kauffman (2009) provide detailed discussions of these and related topics.

150 Chang et al. (2017).

151 As measured with four items adapted from Blackwell et al. (2007).

152 These data require further independent analysis to better understand: (i) if these measures are capturing the desired skills and domains in low- and middle-income country settings, despite being created and calibrated in high-income countries; and (ii) 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 in some instances with improvements in learning outcomes and overall student wellbeing.

Emergency preparedness

COVID-19 has highlighted the importance of preparedness for pandemics and other large-scale emergencies. Adequate preparedness and effective response may help reduce transmission of a pathogen; limit the number of cases, hospitalizations, and deaths; maintain non-outbreak-related essential services; 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 utilized to gauge health system preparedness, as was done for the current COVID-19 pandemic in a previous paper.[153] In addition, SDI surveys have begun adapting 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 as to whether they have an emergency plan in place; whether appropriate contingency mechanisms exist for the isolation, transport, and treatment of contagious patients; and surge capacity.

The COVID-19 pandemic and the associated lockdowns brought a renewed focus on providing continuity in learning when children cannot be physically in school. Students' and teachers' digital skills and 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 UNESCO's UIS 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 further increased 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 tools will need to be further adapted to capture these and other novel aspects of ICT that can shape learning outcomes. With this in mind, the World Bank Group 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 World Bank Group Education Global Practice could allow for additional synergies moving forward. Similarly, to the extent that COVID-19 itself continues to pose a challenge, new physical requirements for spaced seating, WASH infrastructure, and protective equipment become essential elements of school life.

153 Sharma et al. (2020).

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 citizens receive. What factors determine 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 to better understanding why providers make certain good or suboptimal choices in their daily practice and gaining more systematic insight into the role of management.

Better understanding providers' performance

In health, the current SDI vignettes capture whether a provider gives the correct diagnosis and treatment, and are currently being expanded to provide more information on why a provider is diagnosing a condition incorrectly. For example, while the existing data show that many providers are not very effective at diagnosing diabetes, it is not clear whether they think the patient is healthy or instead displaying risk factors for another disease, such as depression or chronic fatigue, that might present similarly to diabetes. Similarly, at present, while health care providers can generally diagnose the TB vignette quite well, 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 non-standard answers given. This information will shed light on the most common errors and will help provide more pinpointed recommendations to improve provider performance.

In schools, teacher observation allows 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, a "snapshot" of current activities in the classroom is noted by the enumerator. The instrument allows an enumerator to easily 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 and/or students are actively engaged in assigned learning tasks. Analysts can further 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 World Bank Group Education Global Practice, was tested for the Mozambique SDI survey in 2018, in place of the standard classroom observation module. TEACH functions in a similar way to the SDI tool but generally involves recording classes on video, then having a trained observer view the videos and provide her/his assessment.[154] In addition to time-on-task, this approach assesses teacher quality practices through behavioral information observed and classified into three categories: classroom culture, instruction, and socio-emotional skills. Recent efforts on this front are trying to leverage machine-learning techniques to automatically identify teacher-student interactions and engagement through data mining of 2-D videos of classroom observations.[155] If proven reliable, these methods might help reduce the cost of data collection and increase the information at researchers' disposal to inform policies that promote effective teaching and learning in classrooms.

154 The class observation can also be 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.

155 See for instance Aung et al. (2018).

The role of management

Limited provider competence and high absenteeism 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 employee motivation and effort.[156] Intrinsic motivation also matters for a provider's performance,[157] together with the financial resources that organizations and staff must work with, and the selection of staff members in the first place.[158] Many of these factors interact with each other.[159] Growing empirical evidence shows that management practices have an important role in how organizations perform.[160]

The explanatory power of management also 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 (though increasing incentives and monitoring among bureaucrats had an opposite effect). Management quality (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, the United States, the United Kingdom, and Uganda.[161] Interventions aimed at improving management quality in schools have also been shown to improve student performance, although the effectiveness of such measures varies by context.[162]

Existing evidence suggests that management practices also affect quality of care in the health sector.[163] Effective health care management requires oversight of many different facets of service delivery, including the facility's layout and patient flow, development and standardization of protocols, staff recruitment and retention, effectively utilizing personnel and equipment, setting targets, continuously tracking performance, and utilizing accountability systems with performance incentives. While 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 of resource constraint. This suggests that, 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 influencing 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 which had been included in Kenya's 2018 SDI survey. The sample includes both public and private facilities, so that differences between management practices and incentive systems by facility ownership can be examined. The survey was designed building off two highly utilized existing management tools, the Development World Management Survey (D-WMS) and the SafeCare Standards, and was adapted to fit the Kenyan primary health care context.[164] Five-hundred facility managers will each participate in an hour-long phone survey, where 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. Not only will this survey allow for benchmarking of management practice outcomes in Kenya against practices in other settings (given that the D-WMS methods are standardized), but this tool will also introduce innovations to measure management practices relating to staff absenteeism and stock outs of equipment, supplies, and medicines.

Management can be equally important for school outcomes. The recent SDI survey in Indonesia was coupled with the implementation of the D-WMS among school directors to assess school management quality. 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

156 See Ashraf, Bandiera, and Jack (2014); Rasul and Rogger (2017); Karachiwalla and Park (2017); Björkman Nyqvist et al. (2018).

157 Ashraf, Bandiera and Lee (2018); Deserranno (2018).

158 Das et al. (2013) and Deserranno (2018).

159 Donato et al. (2017).

160 Bloom, Sadun, and Van Reenen (2016).

161 See Bloom et al. (2015); Di Liberto, Schivardi and Sulis (2015); and Crawfurd (2017).

162 See Blimpo, Evans, and Lahire (2015) and Fryer (2017).

163 Macarayan et al. (2019).

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

school administrators (either principals or their second-in-command), trained and certified enumerators assign scores ranging from 1 to 5 along the five aforementioned areas. Again, preliminary results were included in the 2020 Indonesia SDI report but, due to the richness of the data, its link to quality service provision and student learning will be further analyzed in a separate forthcoming publication. This methodology will continue to be replicated in future SDI surveys whenever possible.

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. While this data has proven to be extremely useful to understand the potential implications of education and health financing reforms at the country level, its collection and cross-country comparability have been challenging due to country-specific financing and organizational structures, as well as data privacy concerns. Likewise, having a broad understanding of expenditure composition (e.g., human resources, equipment, and maintenance, among others) would be invaluable to better understand efficiency differences between facilities and over time. As the SDI program continues to evolve, it will inevitably continue to 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, combined with a reduced 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 otherwise adversely impact 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.[165]

Understanding the individual and family characteristics that are likely to shape children's learning and health are an important complement to understanding service utilization. In education, older surveys only collected 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 education SDI surveys in the Democratic Republic of Congo (forthcoming 2021) and in Indonesia (2019) collected data using household modules from a sub-sample of students tested in SDI schools.[166] These modules include questions on topics that include parental education, household characteristics (e.g., infrastructure, work and sources of income, household structure), assets (both agricultural, such as livestock and crops, and non-agricultural, such as ownership of TV, radio, computer, motor vehicle, or refrigerator), books at home (number and type), and educational expenditures by type (e.g., school fees, food, and others). Likewise, as previously noted, the forthcoming health SDI 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 measurement improvements 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-SD 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 could 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.[167]

Finally, this report offers a preview of how the combination of SDI data and administrative data can help identify bottlenecks in the delivery of health and education services. There may be many reasons behind drug stock-outs 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 improve the overall accountability of the government as a whole, not only the health and education sectors.

165 World Bank (2019).

166 Yarrow et al. (2019).

167 Bau, Das, and Yi Chang (2021).

IMPROVING COMPARABILITY OF SDI SURVEYS OVER TIME

During the past decade, SDI surveys have steadily evolved. Each survey has also been tailored to address specific policy and research questions in its setting of implementation. While this means that the surveys are well designed to respond to pressing national policy concerns, it has also posed challenges for the direct comparability of some surveys from the same country over time. In addition, since the SDI surveys are also designed to provide just-in-time findings in fragile and conflict-affected settings, necessary logistical and safety considerations may also impact 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 Tanzania, Mozambique, and Kenya.[168]

The methods used in the Tanzania education SDI surveys in 2014 and 2016 are an example where survey methods allowed for direct comparisons across the two years. Over this period, student learning outcomes for Grade 4 students, as measured by the English, math, and Swahili SDI tests, improved. The SDI mean student test score across all subjects improved by an average of 11 percentage points. Other SDI indicators in which a significant change was observed in this two-year period included the share of students with pens, pencils, and exercise book (increased from 84 to 92 percent) and the proportion of schools with a functioning blackboard (increased from 74 to 83 percent). Interestingly, the proportion of students with a textbook significantly declined (from 25 to 19 percent), while the observed student-teacher ratio increased (from 43 to 47), perhaps indicating that system resources have not kept pace with enrollment increases. Smaller changes were observed in other SDI indicators, such as teacher classroom absence (declined by 5 percentage points), teachers' subject knowledge in language (slightly deteriorated) and mathematics (increased significantly, but only on certain tasks, such as Venn diagram and graph interpretation). Researchers also found a modest change in pedagogical knowledge (slightly improved). Trends in basic infrastructure showed no noticeable improvement.

The Tanzania 2014 and 2016 education surveys are so far the only SDI with a complete panel of schools tracked over time. The direct data comparability was the result of a careful panel sampling strategy and use of identical survey tools. These methodological and logistical choices allowed 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 substantially improved 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, students' and teachers' performance in SDI assessments may be affected by multiple factors, including how they are feeling 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 carefully 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 education SDI surveys in 2014 and 2018. Over this longer interval, researchers observed significant changes in both student knowledge (increased from 21 to 31 percent of correct answers in a student assessment) and teacher knowledge (increased from 31 to 42 percent of correct answers in a teacher assessment). At the same time, while teacher absence was greatly reduced (from 45 to 30 percent), the percentage of schools with minimum equipment dropped significantly (from 77 to 68 percent). Differences in all the remaining SDI indicators were 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, as 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 fieldwork faced several logistical constraints that prevented enumerators from reaching all schools included in the sample and from carrying out second (unannounced) visits in many others. Moreover, fieldwork for the 2014 survey started approximately three months earlier in the calendar year than did the fieldwork in 2018. This meant that

168 The lessons and statistics in this section are drawn from the analytical results from Trako et al. (2019) for Tanzania, Bassi et al. (2019) for Mozambique, and World Bank and Government of Kenya (2019) for Kenya.

data was collected at different moments of the school year, which could affect: (i) student learning, as students in one survey had more schooling; and (ii) absence indicators, since both teachers and students may be more likely to be absent from school as the school year draws to an end, for example. While the final analysis employed adjusted weights and explored differences using various sensitivity analyses and scenarios, the magnitude of the differences observed could be partially attributed to some of the above circumstances. For instance, one sensitivity analysis suggested that teacher absence for 2018 was between 30 and 40 percent, which would amount to a reduction of anywhere between 5 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 to better understand the role of seasonality in education survey findings.

In Kenya, a health SDI survey was conducted in both 2012 and 2018. As discussed in the 2018 SDI Kenya Survey Report,[169] there are important differences between the 2012 and 2018 survey rounds which may influence observed differences in results. Comparing the 2018 values with those from 2012, Kenya appears to have experienced an increase in caseload, absence rate, and availability of some infrastructure. While the two survey rounds utilized very similar instruments, methodological improvements and differences in the nature of the samples lead to differences in how sampling weights are applied in calculations using data from the two rounds. Because of this, differences between 2012 and 2018 values, particularly for indicators that use provider-level disaggregated data, may reflect differences in weighting schemes, rather than true underlying changes. 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, compared to 2012. The 2012 survey was conducted in 15 counties and is representative only at the national level, while the 2018 round encompassed 47 counties and is representative at both national and county levels. Unlike the 2018 survey, the 2012 survey did not include for-profit private facilities. The challenges in generating comparable results between the two surveys underscore the tradeoff between constantly improving the survey and its methods and ensuring that policy-relevant changes over time can still be examined. Going forward, SDI methods increasingly emphasize gathering and retaining details on survey sampling frames, which can help analysts adjust for sampling differences and improve comparability. Overall, and not unlike other survey

169 World Bank and Government of Kenya (2019).

RESULTS IN ACTION: HOW SDI INFORMS PROGRAM OPERATIONS

In a decade of implementation, the SDI surveys have had significant impact in countries. The surveys serve multiple purposes for policy makers and program implementers. They can help provide insight into aspects of service delivery that are typically not well measured. For known health or education challenges, they can help diagnose root causes. From an operations standpoint, they can help governments and the World Bank measure the success of projects over time. Finally, they can help 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 (DRC), 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 DRC's nascent education management information system, and the SDI data allow oversight by the government in areas that would otherwise be opaque. As a result, the SDI survey informed and strengthen a World Bank's 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 over 1,500 public primary schools and will be repeated again in the same sample of schools in 2022. This baseline and endline will allow the Bank team to assess the project's success in meeting its objectives.

In Madagascar, SDI has been instrumental to offer 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. Following the SDI survey, the Improving Nutrition Outcomes Project was begun, with the goal of increasing the diagnosis and treatment capacities of frontline health care providers. In this case, although the problem of malnutrition was well understood, the SDI survey helped indicate the need for better health care provider training to manage these cases.

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, the government introduced the "Big Results Now" initiative, which aimed to strengthen accountability, improve incentives and provide better support for students. The $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 inform the preparation of the country's Education Sector Plan (ESP) 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 post-COVID: using familiar tools in new ways

In addition to promoting evidence-driven delivery reforms in countries, SDI research contributes to a global measurement practice agenda that is evolving quickly in the wake of COVID-19. COVID-19 is dramatically re-shaping service delivery in many countries and posing new challenges to measurement. Focus in many settings has turned to remote data collection. While 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, un-announced visits to measure teacher absence or the visual verification of medicine stocks in a clinic or of 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 prior to a future in-person SDI survey, generating a panel dataset 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[170] covers the following aspects of essential health services:[171]

- Health workforce (numbers, absences, COVID-19 infections, training, and support);
- Financial management and barriers;
- Service delivery and utilization (facility closures, community communication campaigns, changes in service utilization, and catch-up strategies);
- Infection prevention and control (IPC) capacities (protocols, safety measures, guidelines, and the availability of personal protective equipment (PPE) 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 quickly act. However, such surveys are not a silver bullet. For example, higher rates of non-response, particularly in areas with poor phone connectivity and in facilities already overstretched due to 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 types 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 how they affect the quality of service.

170 This instrument was designed by the WBG Human Development & Public Services team at Development Research Group (DECRG) and the Global Financing Facility for Women, Children and Adolescents (GFF), with the support of the Primary Health Care Performance Initiative (PHCPI) and SDI teams.

171 WHO (2020).

USING MEASUREMENT TO INFORM POLICY DIALOGUE AND STRENGTHEN COUNTRY CAPACITIES

By measuring how services are delivered, results from SDI surveys can help link resources to results and shift national policy dialogue from inputs towards 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 against teacher absence was instituted after the launch of the 2014 SDI Education Report. Togo's SDI education findings sparked a revision of teachers' college curricula and a novel collaboration between a teacher's 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 report 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" program monitoring framework. Policies based on SDI evidence resulted in more equitable deployment of teachers within and across districts in Tanzania, and teachers experienced increased motivation 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 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 further 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 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 received both basic and advanced training in data use and applications.

Future measurement initiatives have the potential not only to reinforce national capacity but also to help 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. 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 re-use 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 above, will allow for deeper analysis and an increased 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, helping add detail where needed. As 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 immediate use of SDI evidence is 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 foundations for inclusive growth. While this is an important contribution, 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 it available and useful to the scientific community, which are harmonized and available through the data portal www.sdindicators.org. The earlier releases of some of these data was 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, more than 15 peer-reviewed academic articles have been published in leading economics, education, and health journals.[172] The use of the data helps underline the surveys' rigor and relevance for both policy and research.

The SDI program expects that this new 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.[173] 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 learnings for every country and can inform many local policy processes. At the same, 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 publishing case studies) can help ensure growing demand for SDI surveys in the years ahead.

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

173 The present report focuses primarily on within- and across-country comparisons. Thus, 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 report.

REFERENCES

Angrist, J.D., and V. Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *The Quarterly Journal of Economics* 114 (2): 533–575. 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: 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.

Attanasio, Orazio P. and Katja M. Kaufmann. 2014. "Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender." *Journal of Development Economics* 109: 203-216.

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, A. Yi Chang. 2021 [Forthcoming]. "New Evidence on Learning Trajectories in a Low-Income Setting." International Journal of Educational Development.

Bergman, P. 2021. "Parent-child information frictions and human capital investment: Evidence from a field experiment." Journal of Political Economy, 129(1), 000-000.

Björkman Nyqvist, M., A. Guariso, J. Svensson, & D. Yanagizawa-Drott. 2019. "Reducing Child Mortality in the Last Mile: Experimental Evidence on Community Health Promoters in Uganda." Forthcoming: *American Economic Journal: Applied Economics.*

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): 246-263. 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." World Bank Policy Research Working Paper No. 7238, World Bank, Washington, DC.

Bloom, N., R. Lemos, R. Sadun, and J. Van Reenen. 2015. "Does management matter in schools?" *The Economic Journal* 125 (584): 647-674.

Bloom, N., R. Sadun, and J. Van Reenen. 2016. "Management as a Technology?" Working paper No. W22327, National Bureau of Economic Research.

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): 262-268.

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." *The Quarterly Journal of Economics* 126(4): 1593-1660.

Crawfurd, L. 2017. "School management and public-private partnerships in Uganda." *Journal of African Economies* 26(5): 539–560, https://doi.org/10.1093/jae/ejx021

Danaei, G., K.G. Andrews, C.R. Sudfeld, G. Fink, D.C. McCoy, E. Peet, A. Sania, M.C. Smith Fawzi, M. Ezzati, and W.W. Fawzi. 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): p.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*: 29-57.

Deserranno, E. (2019). "Financial incentives as signals: experimental evidence from the recruitment of village promoters in Uganda." *American Economic Journal: Applied Economics.*

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." Working paper No. W23437, National Bureau of Economic Research.

Karachiwalla, N., and A. Park. 2017. "Promotion incentives in the public sector: Evidence from Chinese schools." *Journal of Public Economics* 146: 109-128.

Kim, Y. and V. Radoias. 2018. "Screening, diagnosis, and long-term health outcomes in developing countries—The case of hypertension." *PLoS One* 13(12): p.e0208466.

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), pp.1666-1672.

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): 431-437.

Kruk, M.E., M. Myers, S.T. Varpilah, and B.T. Dahn. 2015. "What is a resilient health system? Lessons from Ebola." *The Lancet* 2015;385(9980): 1910–2.

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), p.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): p.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), pp.49-61

Rasul, I., & Rogger, D. (2018). "Management of bureaucrats and public service delivery: Evidence from the Nigerian civil service." *The Economic Journal*, 128(608), 413-446.

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), p.e0171236.

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 No. 9334. World Bank, Washington, DC.

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." United States Agency for International Development, Washington DC.

USAID. 2018. "Madagascar nutrition profile." United States Agency for International Development, Washington, DC.

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

WHO. 2020. "Continuity of essential health services: Facility assessment tool." World Health Organization, Geneva.

WHO. 2020a. "Pulse survey on continuity of essential health services during the COVID-19 pandemic: interim report, 27 August 2020." World Health Organization, Geneva.

WHO. 2020b. "Role of primary care in the COVID-19 response." Manila: WHO Regional Office for the Western Pacific. https://apps.who.int/iris/handle/10665/331921.

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

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

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

Yarrow, N., Afkar, R., Masood, E., Gauthier, B. (2020). "Measuring the Quality of MoRA's Education Services (English)". Washington, D.C.: World Bank Group

CHAPTER 5

LESSONS FROM SDI FOR A POST-COVID-19 WORLD

The COVID-19 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.

The aim of this booklet has been to leverage a decade of SDI data, now fully harmonized across countries, to examine the state of health and education systems, and highlight ways that systems can evolve to better help frontline providers serve patients and students, during this crisis and beyond. The data presented in the report have been collected in a pre-COVID world, thus 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 very systemic weaknesses. This report offers an actionable evidence base for policymakers as they chart a path forward post-COVID-19 and work to make systems stronger and more resilient to future shocks.

RETHINKING SERVICE DELIVERY

On the health side, the surveys show substantial gaps in service quality. Rates of provider absence are high, particularly in public facilities, and the caseload of staff varies 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 focus on supply-chain strengthening, 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 basics.

SDI results suggest margins for strengthening health-service delivery. Health leaders, system coordinators, and facility managers can do more to ensure the presence of health care providers and to fairly balance caseloads by reallocating staff to overburdened facilities. Such efforts may demand politically sensitive tradeoffs between widespread geographical 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 non-physician 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 in reducing infrastructure gaps at rural health facilities.

COVID-19 brings new challenges to health systems. In addition to the urgency of stopping transmission – with both non-pharmaceutical interventions and vaccinations – policymakers need to protect core health service delivery functions and ensure equitable access to care, while managing increased stress on the system, including for critical care.[174] The pandemic-related recession and growing demands on public expenditures will create fiscal pressure on governments. Health spending priorities should be protected in this new environment, including for immediate needs, such as providing COVID-19 diagnostics, surveillance and care, and for longer-term objectives, such as the expansion of universal health coverage. Routine services disrupted by the pandemic will need to resume and, in some cases, catch up significant lags, for example in routine immunization. Reckoning with existing system-level weaknesses will an important step for policymakers 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*. 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 the SDI countries are already performing well, illustrating the potential for each country to bring its students to a high level. Results also underscore the importance of specific school inputs and teacher characteristics that positively correlate with higher student learning. In particular, classrooms with functioning whiteboards; clean, private, and gendered toilets; and more teachers per student are characteristics observed more frequently in schools with better average student test scores. Equally important seems to be having more knowledgeable teachers, with gender and age showing as strong predictors of teachers' effectiveness. Finally, private schools in a selected sub-sample 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 improve students' daily experience in school. Having functioning whiteboards and toilets might seem trivial, but evidence suggests these features can make a big difference for some children. Strong arguments encourage policy makers to 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. Overall, education SDI surveys can help governments to efficiently deploy investments so that more young people can reach their full potential.

174 World Bank (2020e).

Today, COVID-19 threatens educational outcomes in many countries.[175] 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 that have lost access to education. Strategies to remediate schooling losses will require designing and implementing school re-opening protocols adapted to the specificities of the pandemic. At a minimum, these 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.[176] Deeper reforms will be needed in order 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 educational systems, this report 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.[177] To shape resilient education systems, countries will need to draw lessons from worldwide distance-learning experience and expand the infrastructure for online and remote learning. In health, telemedicine can, in some cases, help increase access to higher-quality care in an efficient manner. Finally, in the process of moving towards digital solutions, governments should consider specific disadvantages that some groups face in accessing online platforms, such as gender and disability gaps in access 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 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 can be understood by policy makers and citizens alike. 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.

175 World Bank (2020c).

176 World Bank (2020b).

177 Reimers and Schleicher (2020).

As evidenced in the recent *World Development Report 2021*, the commitment to measurement among policy makers and stakeholders continues to gain force globally. The SDI surveys are part of that agenda. Over the next decade, health and education management information systems (HMIS/EMIS) will likely expand and incorporate a greater range and quality of data, including electronic medical records, student performance data, and more advanced facility monitoring. However, the deployment and/or 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 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. Health and education systems' performance can be measured in multiple ways and, by looking at data from all angles, using diverse instruments, policymakers will be able to more clearly see systems' strengths and weaknesses.

The SDI surveys have expanded geographically in recent years, reflecting the global demand for reliable measurement. SDI efforts in Afghanistan, Armenia, Bhutan, Cameroon, Comoros, Cote d'Ivoire, Democratic Republic of Congo, Guatemala, Guinea Bissau, Indonesia, Iraq, Malawi, Moldova, and Pakistan are ongoing or recently completed during the production of this report. The health surveys in Bhutan, Moldova, and Guatemala have required substantial adaptation and innovation to suit the new contexts. This has meant an expanded focus, with methodological innovations to measure more health conditions, including nutrition, mental health, and non-communicable 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, parental school choice and socioemotional conditions, and socioeconomic and household 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 begin returning to school, and doctors and nurses will hope to resume routine practice. This partial 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 is an opportunity to reinvigorate and transform systems.

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

REFERENCES

Reimers, F., and A. Schleicher. (2020). "Schooling disrupted, schooling rethought: How the Covid-19 pandemic is changing education." Paris: Organisation for Economic Co-operation and Development.

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.

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

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


dacom

APPENDIX

APPENDIX TABLE A1A SDI HEALTH SAMPLE, DISAGGREGATED BY LOCATION, FACILITY LEVEL, AND OWNERSHIP

	All countries		Niger 2015		Mozambique 2014		Sierra Leone 2018		Madagascar 2016	
GDP per capita in survey year			$1,048		$1,328		$1,604		$1,634	
	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Total facilities	7810	-	255	-	195	-	536	-	444	-
Rural	5276	68%	192	75%	172	88%	377	70%	218	49%
Urban	2534	32%	63	25%	23	12%	159	30%	226	51%
Hospitals	872	11%	16	6%	38	19%	30	6%	37	8%
Health clinics	2823	36%	67	26%	157	81%	109	20%	316	71%
Health posts	4115	53%	172	67%		0	397	74%	91	20%
Public	5805	74%	220	86%	193	99%	493	92%	289	65%
Private/NGO	2005	26%	35	14%	2	1%	43	8%	155	35%

	Togo 2013		Uganda 2013		Tanzania 2016		Kenya 2018		Nigeria 2013	
GDP per capita in survey year	$1,761		$2,033		$3,227		$3,461		$5,980	
	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Total facilities	180	-	394	-	383	-	3038	-	2385	-
Rural	126	70%	285	72%	222	58%	2249	74%	1435	60%
Urban	54	30%	109	28%	161	42%	789	26%	950	40%
Hospitals	16	9%	9	2%	30	8%	285	9%	411	17%
Health clinics	46	26%	133	34%	91	24%	594	20%	1458	61%
Health posts	118	66%	252	64%	262	68%	2159	71%	516	22%
Public	143	79%	236	60%	266	69%	1762	58%	2203	92%
Private/NGO	37	21%	158	40%	117	31%	1276	42%	182	8%

Notes: The Nigeria survey included 12 of 36 states due to logistical constraints and the survey in Kenya was representative at the national and county level.

APPENDIX TABLE A1B SDI EDUCATION SAMPLE, DISAGGREGATED BY LOCATION, OWNERSHIP, AND STUDENT BODY SEX

	All countries		Kenya 2012		Madagascar 2016		Mozambique 2014		Niger 2015	
GDP per capita in survey year			$3,461		$1,634		$1,328		$1,048	
	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Total schools	3,297	-	306	-	473	1	-	1	256	-
Rural	2,406	73%	207	68%	319	67%	172	85%	169	66%
Urban	891	27%	99	32%	154	33%	31	15%	87	34%
Public	2,537	77%	239	78%	308	65%	203	100%	223	87%
Private	758	23%	67	22%	165	35%			33	13%
Boys	57	2%	10	3%					3	1%
Girls	40	1%	7	2%					1	0%
Co-ed	3,199	97%	289	94%	473	100%	203	100%	252	98%

	Nigeria 2013		Tanzania 2016		Togo 2013		Uganda 2013		Morocco 2016	
GDP per capita in survey year	$5,980		$3,227		$1,761		$2,033		$8,587	
	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Total schools	760	-	400	-	200	-	400	-	299	-
Rural	604	79%	314	79%	136	68%	325	81%	160	54%
Urban	156	21%	86	22%	64	32%	75	19%	139	46%
Public	458	60%	397	99%	147	74%	319	80%	243	81%
Private	300	39%	3	1%	53	27%	81	20%	56	19%
Boys	27	4%	1	0%			16	4%		
Girls	21	3%	6	2%	1	1%	4	1%		
Co-ed	712	94%	393	98%	198	99%	380	95%	299	100%

Notes: The Nigeria survey includes four states: Anambra, Bauchi, Ekiti, and Niger. For the Kenya survey, three counties of North Eastern province were excluded due to current security concerns.

APPENDIX TABLE A2 DEFINITIONS OF ALL HEALTH AND EDUCATION INDICATORS

HEALTH INDICATORS

1. Infrastructure availability

Availability of an improved water source, an improved toilet, and electricity.	Infrastructure availability is calculated as the availability of three components: improved water source, improved toilet, and electricity. Credit is given if all three components are available. **Improved toilet:** Credit is given if facility reports and enumerator confirms facility has one or more functioning flush toilets or ventilated improved pit (VIP) latrines, or covered pit latrine (with slab). **Improved water source:** Credit is given if facility reports their 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.[178] **Electricity:** Credit is given if facility reports using electric power grid, fuel-operated generator, battery-operated generator, or a solar powered system as their main source of electricity.

2. Medicine availability

Percentage of 14 basic medicines which were available and in-stock at the time of the survey.	Medicine availability is calculated as the percentage of 14 medicines available and in-stock at the time of the survey. The list of medicines included for the SDI is based on a subset of the WHO Essential Medicines list.[179] The medicines included are: 1) Amitriptyline (anti-depressant) 2) Amoxicillin (antibiotic) 3) Atenolol (beta blocker) 4) Captopril (ACE inhibitor) 5) Ceftriaxone (antibiotic) 6) Ciprofloxacin (antibiotic) 7) Cotrimoxazole (antibiotic) 8) Diazepam (anti-seizure) 9) Diclofenac (nonsteroidal anti-inflammatory) 10) Glibenclamide (anti-diabetic) 11) Omeprazole (proton pump inhibitor) 12) Paracetamol (analgesic) 13) Salbutamol (bronchodilator) 14) Simvastatin (statin) The list of medicines in the SDI is adapted based on country standards, and some of these medicines were not included in the surveys in Kenya, Nigeria, and Uganda, so these countries have been omitted from this indicator.

178 WHO and UNICEF (2019). "Water, Sanitation and Hygiene in Health Care Facilities: Global Baseline Report 2019." Geneva: WHO and UNICEF.

179 WHO (2017). "Model List of Essential Medicines." Geneva: WHO.

3. Equipment availability	
Availability and functioning of thermometer, stethoscope, sphygmomanometer and weighing scale.	Equipment availability is calculated as the availability and functioning of a thermometer, a stethoscope, a sphygmomanometer and a weighing scale (adult, child or infant). Credit is given if all four components are available. Thermometer: Credit is given if a facility reports and the enumerator observes that the facility has one or more functioning thermometers (used for measuring patient body temperature). Stethoscope: Credit is given if a facility reports and the enumerator observes that the facility has one or more functioning stethoscopes. Sphygmomanometer: Credit is given if a facility reports and the enumerator observes that the facility has one or more functioning sphygmomanometers. Weighing Scale: Credit is given if a facility reports and the enumerator observes that the facility has one or more functioning adult, child or infant weighing scale.
4. 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 that facility's absenteeism on an unannounced visit is found to be 40%, then the number of health care providers will be adjusted down by 40% and only 6 health care providers will be counted as available for patient care.
5. 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 share of ten randomly sampled workers who should be on duty. Health professionals doing outreach are counted as present.
6. Diagnostic accuracy	
Percentage of correct diagnoses provided in the five clinical vignettes.	The SDI includes five core vignettes: (i) acute diarrhea w/ dehydration; (ii) pneumonia; (iii) diabetes mellitus; (iv) pulmonary tuberculosis; (v) malaria w/ anemia. Health care providers are scored on their ability to provide correct diagnosis on each of those vignettes and their overall score is calculated as the percent of vignettes answered correctly. Further details on diagnostic accuracy for each disease are provided in **Appendix Table A3**.

7. Treatment accuracy	
Percentage of correct treatments provided in the five clinical vignettes.	The SDI includes five core vignettes: (i) acute diarrhea w/ dehydration; (ii) pneumonia; (iii) diabetes mellitus; (iv) pulmonary tuberculosis; (v) malaria w/ 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 percent of vignettes answered correctly. Further details on treatment accuracy for each disease are provided in **Appendix Table A4**.
8. 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 post-partum hemorrhage and seven specific actions for neonatal asphyxia.
9. Inappropriate antibiotic usage	
Percentage of providers inappropriately prescribing antibiotics.	The SDI surveys also collect information on inappropriate antibiotic use, defined as providers that prescribed an antibiotic during the tuberculosis vignettes (aside from the antibiotics recommended as part of the tuberculosis regimen) or any antibiotics for the diarrhea vignettes (for which antibiotics are not indicated given the patient examination). Inappropriate antibiotic usage is calculated as the percentage of health care providers that inappropriately prescribe antibiotics among all health care providers given the clinical vignettes.

EDUCATION INDICATORS

1. School absence rate

Share of a maximum of 10 randomly selected teachers absent from school during an unannounced visit	This indicator is measured as the share of teachers who are absent from school at the time of an unannounced visit. It 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 (excludes volunteer and part-time teachers) who are on the school roster. The whereabouts of these 10 teachers are then verified in the second, unannounced, visit. Teachers found anywhere on the school premises are marked as present.

2. Classroom absence rate

Share of teachers who are present in the classroom during scheduled teaching hours as observed during an unannounced visit	The indicator is measured as the share of teachers not in the classroom at the time of an unannounced visit. The indicator is constructed in the same way as the school absence rate indicator, except that the numerator now is the number of teachers who are either absent from school, or present at school but absent from the classroom.

3. Time spent teaching per day

Amount of time a teacher spends teaching during a school day	This indicator reflects the typical time that teachers spend teaching on an average day. It combines data from the staff roster module (used to measure absence rate), the classroom observation module, and reported teaching hours. The teaching time is adjusted for the time teachers are absent from the classroom, on average, and for the time the teacher teaches while in classrooms, based on classroom observations. While inside the classroom, distinction is made between teaching and non-teaching 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 working on a specific task, drilling, or memorization. Non-teaching activities includes working on private matters, maintaining discipline in class, or doing nothing, and thus leaving students not paying attention.

4. Minimum knowledge

Share of teachers with minimum knowledge	This indicator is measured as the percentage of teachers who can master the curriculum they taught. It is based on mathematics and language tests covering the primary curriculum administered at the school and is calculated as the percentage of teacher who score more than 80 percent on the language and mathematics portion of the test. The test is given to all mathematics or language teachers who taught third grade last year or fourth grade in the year the survey is conducted.
Test score	It is measured as the overall score of mathematics, language, and pedagogy tests covering the primary curriculum administered at the school level to all mathematics and language teachers who taught third grade last year or fourth grade in the year the survey was conducted.

5. Minimum infrastructure availability	
Unweighted average of the proportion of schools with the following available: functioning electricity and sanitation	It is a binary indicator capturing availability of (a) functioning toilets and (b) classroom visibility. Functioning toilets is defined as whether toilets were 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.

6. 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	It is a binary indicator capturing availability of (a) functioning blackboard and chalk and (b) pens, pencils, and exercise books in fourth-grade classrooms. In one randomly selected fourth-grade classroom in the school, the enumerator assessed if there was a functioning blackboard by looking at whether text written on the blackboard could be read at the front and back of the classroom, and whether there was chalk 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%.

7. Share of pupils with textbooks	
Number of mathematics and language books used in a grade four classroom divided by the number of pupils present in the classroom	The 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 randomly selected class is observed) in one randomly selected fourth-grade class and divided by the number of students in that classroom.

8. Observed pupil-teacher ratio	
Average number of grade-four pupils per grade-four teacher	This indicator reflects the typical ratio of pupils to teachers in the fourth-grade classroom. It is measured as the number of students in one randomly selected fourth grade class at the school.

APPENDIX TABLE A3 DEFINING A CORRECT DIAGNOSIS

DISEASE	NOTES
Diarrhea with dehydration	The Integrated Management of Childhood Illnesses (IMCI) guidelines[180] 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 would be suggested by IMCI guidelines. This includes 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. Countries varied in how they asked about diagnosis, and some countries listed diagnoses jointly. Nigeria and Uganda listed "acute diarrhea with severe dehydration" as the only diagnosis possible. Madagascar listed "diarrhea with moderate dehydration" and "diarrhea with severe dehydration." Togo listed "diarrhea with severe dehydration." All of these answers were counted as correct. Note that for these countries we are not able to provide an accurate estimate of how many doctors diagnosed diarrhea alone, we can only 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, clinicians should arrive at a diagnosis of diabetes type II as 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 has also been classified as correct. This 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 were counted as correct. All countries also included "severe malaria" as an option and this was not counted as correct since the case does not meet the definition of severe, and severe malaria would require different treatment. Anemia was listed simply as "anemia" without a severity specified. Providers therefore received credit for specifying malaria or simple malaria AND anemia. Kenya excluded this module entirely so is omitted and their diagnostic accuracy is counted as the average of the four other vignettes.
Post-partum hemorrhage	A diagnosis of post-partum hemorrhage is counted as the correct response.
Neonatal asphyxia	A diagnosis of neonatal asphyxia or respiratory distress is counted as the correct response.

180 World Health Organization. "IMCI Chart Booklet." World Health Organization, Integrated Management of Childhood Illnesses, 17 Sept. 2017, www.who.int/maternal_child_adolescent/documents/IMCI_chartbooklet/en/.

APPENDIX TABLE A4 DEFINING A CORRECT TREATMENT

DISEASE	NOTES
Diarrhea with dehydration	The Integrated Management of Childhood Illnesses (IMCI) guidelines note that correct treatment of diarrhea with severe dehydration would be to give intravenous fluid immediately, to give an NG tube if that is not possible and to refer to a higher-level facility if both treatments are unavailable. If the child has only some dehydration, then ORS is the recommended treatment. Given the symptoms here the correct treatment should be rehydration with an IV or NG tube. However, the use of ORS plus zinc was also counted as correct. Since the child was able to drink in most cases, the providers may have incorrectly believed that this was a less severe case of dehydration. Correct treatment rates would be much lower if only IV fluids/NG tube were counted as correct (13% correct). Uganda did not include an option for ORS so only treatment with IV fluids/NG tube was taken as a correct response. Kenya did not include an option for IV fluids/NG tube so only ORS with zinc was 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 and IMCI guidelines recommend an anti-pyretic in this case. Correct treatment was counted as treatment with amoxicillin, a first-line antibiotic, and any anti-pyretic.
Diabetes	The World Health Organization's (WHO) Package of Essential Noncommunicable Disease Interventions (PEN) protocols[181] state that "individuals with persistent fasting blood glucose >6 mmol/l despite diet control should be given metformin and/or insulin as appropriate." For our purposes, correct treatment was counted as any hypoglycemic (including insulin) or referral to specialist. Referral to a higher level was the recommended protocol for diabetes at the primary level in multiple countries so that option was counted as correct. Although PEN protocol would suggest diet control before prescribing hypoglycemics, we have allowed for prescriptions upon first presentation to count as correct.
Tuberculosis	The WHO Guidelines for Treatment of Tuberculosis[182] recommend combination therapy, ideally with fixed dose combination (FDC). Providers were simply required to mention combination therapy. Knowledge of correct duration and dosage was not necessary and providers would score worse if those were required. For example, 23% of providers prescribe 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 were assessed on having gotten the correct dosage/timing. This therefore 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.

181 World Health Organization. "WHO PEN - Package of Essential Noncommunicable Disease Interventions." Noncommunicable Diseases and Their Risk Factors, *World Health Organization,* 22 July 2019, www.who.int/ncds/management/pen_tools/en/.

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

Malaria with anemia	IMCI guidelines recommended 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 was given for treatment with any artemisinin combination therapy (ACTs) or artemether-lumefantrin (coartem). In addition to anti-malarials, the provider must prescribe paracetamol and iron for the anemia. Kenya did not include the malaria vignette and was excluded. Nigeria and Uganda did not include questions about iron and were excluded for the sake of comparability.
Post-partum hemorrhage (PPH)	The WHO Recommendations for the Prevention and Treatment of Postpartum Hemorrhage[183] state that "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." There are five specific actions that a provider should take. These were: 1) Proposed to determine cause 2) Proposed any uterotonics 3) Proposed an IV line 4) Proposed a foley catheter 5) Proposed bimanual uterine massage. Uterotonics included oxytocin or another uterotonic drug, such as misoprostol, ergometrine or a prostaglandin. Togo and Uganda only gave credit to health care providers if they were able to determine the correct dosage of oxytocin.
Neonatal asphyxia	The WHO Guidelines on Basic Newborn Resuscitation[184] provide detailed recommendations on newborn care and actions which should be taken if the child is exhibiting danger signs. Assessments for neonatal asphyxia varied between countries. There were eight actions which were assessed in all countries and these were used as our standard of assessment: 1) Call for help 2) Dry baby 3) Keep baby warm 4) Check to see if baby is breathing 5) Place baby in a natural position 6) Initiate resuscitation with bag/mask 7) Check heartrate This set of actions does not represent the full standard of treatment for neonatal asphyxia but does allow for a common set of actions upon which health care providers can be compared across our sample of countries. Correct treatment is counted as providing at least half of these actions.

183 World Health Organization. "Recommendations for the Prevention and Treatment of Postpartum Haemorrhage," Sexual and Reproductive Health, *World Health Organization,* 6 Feb. 2019, www.who.int/reproductivehealth/publications/maternal_perinatal_health/9789241548502/en/.

184 World Health Organization. "Guidelines on Basic Newborn Resuscitation." Maternal, Newborn, Child and Adolescent Health, *World Health Organization,* 8 Feb. 2019, www.who.int/maternal_child_adolescent/documents/basic_newborn_resuscitation/en/.

APPENDIX TABLE A5 HOW DO HIGH- AND LOW-PERFORMING SCHOOLS DIFFER IN INPUTS AND TEACHER TRAITS?

VARIABLE	N	(1) BOTTOM MEAN/SE	N	(2) TOP MEAN/SE	N	(3) ALL MEAN/SE	T-TEST DIFFERENCE
Functioning toilets (available + clean + private + gendered)	127	0.150 [0.032]	126	0.468 [0.045]	2537	0.381 [0.010]	-0.319***
Visibility indicator (enumerator - front of the classroom)	127	0.858 [0.031]	125	0.920 [0.024]	2513	0.888 [0.006]	-0.062
Share of observed students with pen/pencils	127	0.811 [0.026]	124	0.938 [0.019]	2510	0.933 [0.004]	-0.127***
Share of observed students with exercise book	127	0.752 [0.031]	126	0.910 [0.024]	2518	0.884 [0.005]	-0.159***
Share of observed students with textbook	127	0.227 [0.031]	126	0.427 [0.038]	2509	0.356 [0.008]	-0.200***
Board functioning indicator	126	0.730 [0.040]	126	0.937 [0.022]	2531	0.829 [0.007]	-0.206***
Multi-grade classes indicator	127	0.291 [0.040]	126	0.175 [0.034]	2531	0.213 [0.008]	0.117**
Pupil-teacher ratio	127	56.055 [3.827]	126	34.683 [1.787]	2535	40.280 [0.486]	21.371***

VARIABLE	N	(1) BOTTOM MEAN/SE	N	(2) TOP MEAN/SE	N	(3) ALL MEAN/SE	T-TEST DIFFERENCE (1)-(2)
Mean Teacher Test Score (Task 2 only) [IRT MLE]	126	-1.403 [0.134]	126	-0.019 [0.091]	2274	-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]	2274	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]	2265	-0.188 [0.013]	-0.644***
Mean teacher age	127	35.323 [0.616]	126	39.524 [0.671]	2280	37.710 [0.133]	-4.201***
Proportion of male teachers	127	0.797 [0.021]	126	0.402 [0.024]	2281	0.529 [0.006]	0.396***
Proportion of teachers with less than primary education	127	0.265 [0.035]	126	0.126 [0.022]	2281	0.194 [0.006]	0.138***
Proportion of teachers with secondary education	127	0.290 [0.028]	126	0.350 [0.028]	2281	0.373 [0.007]	-0.060
Proportion of teachers with above secondary education	127	0.365 [0.037]	126	0.322 [0.031]	2281	0.294 [0.007]	0.042
Proportion of teachers with any certification	126	0.820 [0.023]	126	0.652 [0.031]	2278	0.731 [0.006]	0.168***
Share of time teaching	126	0.810 [0.029]	126	0.751 [0.024]	2278	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]	2231	0.360 [0.006]	0.066

The value displayed for t-tests are the differences in the means across the groups.

***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

APPENDIX TABLE A6 EXPLAINING VARIATION IN STUDENT TEST SCORES

TEST SCORES ON SCHOOL INPUTS	(1)	(2)	(3)	(4)	(5)
Board Functioning Indicator	0.179*** (0.050)	0.174*** (0.050)	0.148*** (0.050)	0.151*** (0.050)	0.142** (0.060)
Toilets available + private + gendered		0.087* (0.045)	0.073 (0.045)	0.073 (0.045)	0.060 (0.052)
Pupil-Teacher ratio			-0.006*** (0.001)	-0.006*** (0.001)	-0.005*** (0.001)
Teacher Characteristics PCA	0.231*** (0.026)	0.227*** (0.026)	0.208*** (0.026)	0.200*** (0.026)	0.221*** (0.030)
Dummy equals 1 if school is in urban area	0.243*** (0.039)	0.233*** (0.039)	0.217*** (0.039)	0.155*** (0.044)	0.111** (0.055)
Travel time category: 30-60 min				-0.213** (0.088)	-0.162* (0.096)
Travel time category: 1-2 hr				-0.268*** (0.076)	-0.203** (0.095)
Travel time category: 2-4 hr				-0.257*** (0.074)	-0.182** (0.090)
Travel time category: 4+ hr				-0.225*** (0.073)	-0.126 (0.094)
Poverty Proxy (ln Night Lights)					0.013 (0.010)
Constant	-0.507*** (0.045)	-0.548*** (0.049)	-0.283*** (0.066)	-0.069 (0.087)	-0.126 (0.096)
Adjusted R^2	0.390	0.391	0.403	0.406	0.408
Observations	2453	2453	2453	2453	1979

Country-fixed effects included in all specifications and robust errors are shown

* p<0.10, ** p<0.05, *** p<0.01

TECHNICAL APPENDICES

TECHNICAL APPENDIX 1: SAMPLING PROCEDURES

The sampling strategy for SDI surveys is designed towards attaining indicators that are accurate and representative at the national level to allow for proper cross-country benchmarking and over-time comparisons, when applicable.

In addition, other levels of representativeness are sought to allow for further disaggregation during the analysis stage. Often, samples are constructed to be representative of sub-regions (regions or provinces), rural and urban areas, public and private providers, and other relevant features (for example, facility type, performance, or others). This requires some level of stratification,[185] 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 adequately capture data. Power calculations using pre-survey data are computed to estimate required sample sizes and minimum detectable effect sizes (MDE).

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 (i.e. within some desired confidence intervals) on grade-four pupils' performance 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. After defining the sampling frame (i.e., the complete list from which sampling units are drawn)[186] 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 geographical locations (e.g., districts, communities, counties, neighborhoods) so as 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 (PPS) of the cluster (measured

FIGURE TA1.1 TYPICAL SAMPLING STRATEGY PROCESS FOR SDI SURVEYS

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

186 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 by using additional sources in a data cleaning process to avoid missing entries, duplicates, and outdated or incorrect information.

by the location's number of facilities, providers, or pupils), which helps 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 PPS) as secondary sampling units.[187] At a third stage, a fixed number of health and education workers and pupils are 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 due to 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 back-up facilities are typically not allowed to be used for logistical ease.[188]

Because of the sampling process described above, 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. Note that different weights will need to be applied depending on the relevant level for the estimate which can be the facility, the staff/provider, or the pupil.[189] These different weights are later included in the datasets to facilitate re-estimations.

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 during the Education SDI surveys in Niger (2015). Likewise, in Mozambique (2014) and Tanzania (2016), private schools represented 3% and 1% of the schooling system, respectively, and were therefore excluded for practical reasons. For the 2013 Health SDI in Nigeria, tertiary facilities were excluded from the surveys, as there were very few of them (11 facilities across 6 states), and they would not be appropriately surveyed 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 (adhering to curricula distinct to the national curriculum) were excluded from the 2013 Education SDI surveys in Togo. For the 2015 Education SDI in Niger, *écoles franco-arabes* were not considered (they followed a Koranic curriculum). For the 2014 Health SDI in Tanzania, 93 facilities (approximately 1.2% of the total universe of health facilities) were deleted from the frame because they corresponded to regional hospitals, dental clinics, specialized clinics, or facilities which served prison populations.

Likewise, locations can be discarded from the sampling process due to security concerns (e.g., 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 Education SDI due to security concerns. Similarly, in Niger in 2015, schools from the region of Diffa, which was in a state of emergency, were excluded from the Education SDI surveys. For the 2014 Health SDI surveys in Mozambique, the sample was reduced from 300 health facilities originally to 204 facilities, due to logistical and financial problems. However, these exclusions did not affect the representativeness of results at the national level.

187 In some cases, this second stage becomes the first stage of sampling when schools or facilities are directly drawn from each defined stratum, skipping the selection of cluster locations as a first step (e.g., Niger's 2015 Education SDI).

188 Steps to generate back-ups units for surveyed individuals (e.g., staff and pupils) are also planned during sampling. For instance, for the 2015 Education SDI survey in Niger, a risk of pupils not returning after lunch was identified during pre-testing. 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 that were kept in reserve in case one of the 10 original sampled pupils was not available.

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

There have been very few exceptions when, due to 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 due to the size of the country. The surveys are, nonetheless, representative for each of the selected states.

Finally, some locations or facilities are favored for selection when they are deemed unique in the sample or their characteristics represent opportunities for interesting comparisons or analysis. For instance, for the 2012 Education SDI 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), Nyamira, and Siaya (both as interesting comparison cases[190]). For the 2018 Health SDI surveys in Sierra Leone, all hospitals and health centers were selected within stratum, while other facility types 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 non-randomly select or eliminate certain geographic locations or facility types are taken ex ante and in a fashion that does not compromise rigor and representativeness of results or comparability within and across countries. This ensures that all indicators have a known and acceptable margin of error and can be used for monitoring and evaluation purposes.

TECHNICAL APPENDIX 2: SURVEY METHODOLOGY

The goal of the SDI initiative is to provide national and sub-national evidence on the quality of primary education and basic health service delivery from the perspective of an average citizen who accesses these services. To do so, SDI survey instruments are meticulously structured to collect information on three aspects of service delivery: a) providers' effort, b) providers' knowledge, and c) facility inputs. Provider effort is measured by collecting data on teacher and health-worker 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 the 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, health-facility 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.[191] SDI survey instruments are nimble and use cutting-edge data collection methods, allowing for relatively rapid fieldwork and data analysis. In addition, SDI projects involves 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 fieldwork is randomly checked by trained supervisors. SDI analysis follows internationally accepted indicator definitions and uses reproducible statistical programming.

The overall duration of the SDI project depends on the sample size and country context. As a rough estimate, an SDI survey can take between 12 to 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 SDI with the country's sector strategy. These initial conversations involve virtual and in-country consultations and, potentially, field visits to facilities.

190 Both counties were located in Nyanza province, relatively rural, and with high poverty rates. However, both also showed very different educational performances in terms of passing grades and thus presented good cases for a comparison analysis.

191 SDI facility surveys are often coupled with household surveys to provide a complete perspective on basic health and primary school service quality.

After the SDI survey's overall goals are defined and agreed, the SDI Core Team, Task Teams, and a local technical working group collaborate to contextualize and adapt the survey instruments[192] and define sampling procedures. The local technical group includes ministry and regional officials, development partners, and colleagues from research/academic institutions. Survey protocols and instruments are submitted to the corresponding authorized Ethics Committee or Institutional Review Board for approval. Once surveys instruments are finalized and approved, a Computer-Assisted Personal Interviewing (CAPI) 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: a) standardized materials and suggestions for survey adaptation, b) technical guidance on survey design and sampling procedures, c) protocol for and assistance in quality control, and d) 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.

Government authorities, who constitute the clients of this project, 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 sub-national stakeholders.

Finally, the fieldwork team (Field Coordinator, Survey Firm, and IT technicians) agree on a contract to fully comply with survey protocols and guidelines when they conduct 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 upon 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 2-3 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. He/she is responsible for overall data quality control during the implementation stage.

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

The responsibilities of the enumerators include completing all required interviews, observations, and assessments/tests. Enumerators are also expected to check completed survey forms to ensure that all questions were asked and responses were properly recorded. This 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 tablets and software malfunctions during the fieldwork.[193]

Before the field work starts, introduction letters are sent to all sampled facilities to ensure 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 prior to the arrival of the field team. The letters also specify the number of visits that field teams will make to the sampled facilities.[194] Once

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

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

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

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 (e.g., ministry) to introduce the survey.

Data collection activities consist of two visits to each sampled facility: a first, announced visit and a second, unannounced visit (an example on how survey visits are usually structured is shown in Figure TA2.1).

During the announced visit, enumerators collect data on:

a) The facility's inputs and procedures;

b) Full roster of teachers/health workers and their qualifications;

c) Professional knowledge of up to 10 randomly selected teachers (measured using 4th grade level tests and pedagogical assessments) or health workers (measured using patient case simulations on symptoms of high-burden diseases);

d) Knowledge of up to 10 randomly selected pupils in math, language, and non-verbal reasoning (in schools);

e) Classroom observation for 45-60 minutes of 1 randomly selected lesson (only for education SDIs);

f) Perceptions and satisfaction of visiting clients (only in health facilities and for some countries).

During the second, unannounced visit, which is at least five days apart from the first, announced visit, enumerators collect data on:

a) Attendance check of pre-sampled teachers/health facility workers;

b) Unstaffed classrooms (only for schools).[195]

FIGURE TA2.1 EXAMPLE OF A TYPICAL SDI EDUCATION SURVEY INSTRUMENT (KENYA 2013)

Module #	Module Title	Interviewee	Description
Module 1	School Information	Principal / Head Teacher	Administered to the head of the school to collect information about school type, facilities, school governance, pupil numbers and school hours.
Module 2	Teacher Absence & Info	2A: Head Teacher 2B: Selected teachers	Administered to head teacher and individual teachers to obtain a list of all school teachers, to measure teacher absence and to collect information about teacher characteristics.
Module 3	School Finances	Principal / Head Teacher	Administered to the head teacher to collect information about school finances.
Module 4	Classroom Observation	Observation of teachers and pupils	An observation module to assess teaching activities and classroom conditions.
Module 5	Pupil Assessment	Pupils	A test of pupils to have a measure of pupil learning outcomes in mathematics and language in grade four.
Module 6	Teacher Assessment	Teachers	A test of teachers covering mathematics and Language subject knowledge and teaching skills.

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

Detailed coordination and specialized skills are required to successfully implement SDI surveys. Therefore, survey firms comply with strict selection and training processes when they hire enumerators and constitute 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 successfully conduct this section of the survey, 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 which must be integrally followed during fieldwork 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 they 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 then silently takes note of the clinical questions and procedures provided by the health worker during the simulation exercise, while the Patient Enumerator provides the associated standardized response to each of the health worker's clinical question and procedures.

Distinct tasks assignments are also distributed to enumerators during SDI survey school visits. For instance, during the first, announced visit, one enumerator could interview the principal/head teacher to complete the school information module (module 1), collect the list of all teachers (module 2A), and complete the school finances module (module 3); while another enumerator could observe class lessons (module 4), and administer pupils and teachers' tests (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 non-response. 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. As 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/her.

A second level of quality control is then implemented by the Field Coordinator. The Field Coordinator often uses 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 recommendations 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 analysis to ensure smooth conduct of survey activities and recommend timely rectifications, if needed.

During the third stage, Analysis and Capacity Building, two-week workshops are conducted to build the technical working group's capacity around data analysis and interpretation, undertaking preliminary analyses of core indicators. Key results tables and summary results slides 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 help develop policies and design interventions that can improve health and education outcomes in the client country and around the globe. High public awareness of SDI results mobilizes policymakers 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 FY2020-21, SDI survey work will advance in several countries in Africa and other global regions. Participating countries are at different stages of the SDI process, ranging from procurement to data collection and analysis.

TECHNICAL APPENDIX 3: METHODOLOGICAL GROUNDWORK FOR THE SDI TEACHER AND STUDENT ASSESSMENTS

The 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 et al. 2012). This study reviewed, analyzed and summarized the national Grade 3 and 4 curricula of 13 countries in order to identify common themes (e.g., problem solving, long division). The 13 countries whose curricula were studied were Botswana, Ethiopia, Gambia, Kenya, Madagascar, Namibia, Nigeria, Rwanda, 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 teacher and student assessment 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 the content is based on material drawn from the national curricula of the 13 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. Due to 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, lunch break or after classes). Unfortunately, only teachers that are present in the school can be assessed, so the results represent an upper boundary if we assume that teacher absence and knowledge are negatively correlated.

The SDI student assessment includes three sections: mathematics, language and non-verbal 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 sub-sample 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 non-verbal 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, addition, subtraction, and multiplication up to three digits, divisions up to two digits, fractions, a word problem needing arithmetic, and completing a sequence of numbers. Because of the multiple languages tested and the small number of non-verbal 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 as a large-scale international assessment. The content of the student assessment was also derived from Cunningham et al. (2012) and included topics covered in the Grades 3 and 4 national curricula of the 13 selected 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 non-verbal 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/or missing information from students who are unable to read. For the written component, the proctor informs the student that he/she can answer questions in any order.

Also in contrast with the teacher assessment, the time of the year in which the SDI student assessment is implemented might affect results, particularly since the items are drawn from the Grade 3 and 4 curricula and the assessment is taken by students somewhere in the middle of the Grade 4 schooling year. While this is important for

interpretation of observed differences, SDI students' assessments are not the only international assessment that suffers 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 a number of 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 (HLO) team to have SDI assessments results included in this database, in instances where it is useful and no better information is available. This effort will help increase the precision and availability of other World Bank Group initiatives such as the Human Capital Index (HCI).

REFERENCES

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

Roberta Gatti is the World Bank's Chief Economist of the Middle East and North Africa (MENA) region, and former Chief Economist of the Human Development practice group, where she led the Service Delivery Indicators (SDI) and the Human Capital Index (HCI) initiatives. She joined the WBG 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 also authored multiple WBG flagship reports, including "Jobs for Shared Prosperity" and "Being Fair, Faring Better". Roberta holds a BA from Università Bocconi and a Ph.D. in Economics from Harvard University. She has taught at Georgetown and Johns Hopkins Universities.

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. Kathryn was previously a Young Professional in the Human Development Chief Economist's Office (HDCEO), where she supported SDI survey design and implementation. She holds an ScD in Global Health and Population from the Harvard T. H. Chan School of Public Health, and previously worked at the World Health Organization and Institute for Health Metrics and Evaluation.

Ciro Avitabile is a Senior Economist in the Latin America and Caribbean Education global practice, where he works on supporting investments for human capital development based in Lima, Peru. Prior to joining the WBG, he was an Economist at the InterAmerican Development Bank and received his PhD in Economics at University College London.

Ruben Conner is an Extended Term Consultant in the HDCEO, where he has guided SDI surveys in Armenia, Libya and Moldova. Prior to joining the Bank, he did outbreak response for Doctors Without Borders in South Sudan and worked on malaria programs at PATH. His areas of focus include costs and efficiency of health systems, large-scale impact evaluations and vaccination planning. He is currently pursuing his Doctorate in Public Health (DrPH) at Johns Hopkins University.

Jigyasa Sharma is a Young Professional in the Health, Nutrition, and Population Global Unit, supporting and leading analytical work related to health system strengthening and resilience. Previously in the HDCEO, she provided technical assistance in the design and implementation of the SDI surveys. She holds a ScD in Global Health and Population from the Harvard T. H. Chan School of Public Health.

Andres Yi Chang is an Economist in HDCEO, where he also serves as the SDI Program Manager. He began working at the WBG in 2015, first at the World Bank Research Development Group and then moving to the HDCEO in 2019. Prior to joining the WBG, 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 Latin America, Africa, and Southeast Asia.

WORLD BANK GROUP