

POLICY RESEARCH WORKING PAPER

5690

Making Services Work

Indicators, Assessments, and Benchmarking
of the Quality and Governance of Public Service Delivery
in the Human Development Sectors

Ariel Fiszbein

Dena Ringold

F. Halsey Rogers

The World Bank
Human Development Network
Office of the Chief Economist
June 2011



Abstract

Improving governance is central to improving results in human development. It is clear that money is not enough: improved outcomes from service delivery require better governance, including mechanisms for holding service providers accountable and appropriate incentives for performance. There is therefore a growing demand for indicators to measure how and whether these processes work, and how they affect health and education results.

This paper makes the case for measuring governance

policies and performance, and the quality of service delivery in health and education. It develops a framework for selecting and measuring a set of indicators and proposes options, drawing from new and innovative measurement tools and approaches. The paper proposes the adoption of a more systematic approach that will both facilitate the work of health and education policymakers and allow for cross-country comparisons and benchmarking.

This paper is a product of the Office of the Chief Economist, Human Development Network. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at afiszbein@worldbank.org, dringold@worldbank.org, and hrogers@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Making Services Work

Indicators, Assessments, and Benchmarking of the Quality and Governance of Public Service Delivery in the Human Development Sectors

Ariel Fiszbein, Dena Ringold, and F. Halsey Rogers¹

JEL Codes: H51, H52, I15, I18, I25, I28

Keywords: Governance, service delivery, indicators, public sector performance, education, health, education finance, health care finance, information, teacher motivation and management, health worker motivation and management

¹ Chief Economist, Senior Economist, and Lead Economist, Human Development Network, the World Bank.

Improving governance is central to improving results in human development (HD).² At the country and international level, there is a growing recognition that money is not enough: improved outcomes from service delivery require better governance, including incentives for performance and mechanisms for holding service providers accountable. At the same time, donors and development banks are increasingly in the business of supporting innovative governance reforms through finance and analytical work that aim to strengthen the governance of service delivery. This has led to a growing demand for indicators to measure how and whether these processes work, and how they affect results. Such indicators play several useful roles: they catalyze action by benchmarking service delivery, they inform governments seeking to improve policies and programs, and they track progress.

This paper makes the case for measuring governance and the quality of service delivery in the human development sectors and develops a framework for selecting and applying specific health and education indicators. It proposes the adoption of a more systematic approach that will both facilitate the work of practitioners and allow for cross-country comparisons and benchmarking. The first section of the paper introduces the conceptual framework; the second section discusses a potential “long list” of indicators, and the final section briefly discusses data collection instruments, methodologies, and next steps.

² This paper benefitted from conversations with many colleagues, including Peter Berman, Eduard Bos, Mukesh Chawla, Elizabeth King, Margaret Koziol, Sebastian Martinez, Harry Patrinos, Harry Reid, Emiliana Vegas, Marko Vujcic, Andreas Seiter, and Stephanie Trapnell. Please send comments to afiszbein@worldbank.org, dringold@worldbank.org, and hrogers@worldbank.org.

Why Measure Governance and Service Delivery in Health and Education?

It has now been eight years since the launch of the *World Development Report 2004: Making Services Work for Poor People* (World Bank 2003). The report helped spark a flurry of World Bank operational and analytical work on governance and service delivery in health and education, at a time when academic researchers were also taking a greater interest in the issue. That work has focused on measuring service delivery at the point of contact between provider and client. By identifying gaps in the quality and quantity of “street-level” service delivery, this measurement effort complements the essential work of measuring outcomes such as learning, educational attainment, and health status. At the same time, over the past decade, a shift toward more rigorous evaluation has increased the focus on measuring the impact of health and education interventions that seek to improve service delivery and outcomes by strengthening accountability mechanisms and incentives for providers. In education, these interventions include school-based management, teacher incentives, and the provision of information to users. Similarly, in health, a range of pay-for-performance schemes (and the associated accountability mechanisms) are being evaluated in numerous countries.³

From the perspective of service delivery, governance can be understood as the set of incentives and accountabilities that affect the way provider organizations, their managers, and staffs behave, as well as the quality and efficiency with which they deliver services. From this

³ See http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1239047988859/5995659-1239048041095/Basic_Ed_Cluster_HDNCE.pdf and http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1239047988859/5995659-1239048041095/P4P_in_Health_Cluster_HDNCE.pdf.

vantage point, what is of interest is how providers are selected, paid, monitored, and held accountable for their performance.

Why should we adopt a more structured approach to these HD governance indicators? We see at least three important reasons: to allow more international (and intra-national) benchmarking and spur action, to make impact evaluations more feasible and informative for policy, and to improve monitoring of development projects.

Better and More Standardized Measurement Allows Benchmarking and Spurs Action

An old maxim holds that you cannot improve what you do not measure. Good measurement allows policymakers to see where service delivery and governance are falling short, allowing them to focus on the key problem areas. If the indicators are disseminated to civil society, measurement can help build societal consensus for reform of ineffective governance structures and promote accountability for better service delivery. In the context of specific projects, governance and service delivery indicators constitute key elements of frameworks for monitoring results.

In this process, international, or at least intra-national, comparability can be very helpful. In many of these service delivery areas, measurement without context may not be enough to make it clear whether service delivery is “falling short”. Take, for example, the case of leakage of public funds. If 87 percent of central-government funds intended for schools make it through various layers of government down to the school level, does that suggest good governance or

poor governance? If 25 percent of doctors are absent from rural health centers at the time of surprise visits, does that suggest relatively strong or weak accountability for performance?

If at least a core group of indicators are measured in a standardized way across countries, it is easier to answer these questions. In the case of fund leakage, for example, comparison with the findings of the original Uganda Public Expenditure Tracking Survey (PETS)—where only around 20 percent of cash transfers made it to the school level—might suggest that the country is not doing too badly (Reinikka and Svensson 2005). Ditto with doctor absence, at least compared to the 74 percent absence rate found in small rural clinics in Bangladesh several years ago (Chaudhury and Hammer 2004). These examples are simplistic, but suggestive.

Benchmarking intra-nationally can be even more effective in sparking action, because it allows the states or regions with the most effective service delivery to set the standard for government and the public, while making it harder to make the case that poor service delivery or governance are endemic given local culture or political environment. Two illustrations of how this intra-national comparative approach has been shown to be effective include the case of Papua New Guinea, where the World Bank worked with the Health Metrics Network to strengthen collection of infant mortality data in a way that allowed comparison among parliamentarians' districts, and India, where collection of teacher absence data that was representative at the state level increased media attention to gaps in performance among them (Kremer et al. 2005). For this intra-national comparison purpose, it is necessary to standardize measurement only at the national level, but achieving even this level of coordination and standardization may require conscious attention. Particularly in large federal countries with a

large degree of autonomy at the state or provincial level, there is a risk of state-specific studies that use non-compatible measures of performance. In addition, having international comparators will strengthen the value of this intra-national measurement.

Better Measurement Makes Impact Evaluation Possible

Ultimately, we want to know not only how effective governance and service delivery are, but also what to do to improve them. This requires evaluating the impact of different interventions, at each stage along the results chain. Good measurement of governance and service delivery are essential to tracing these impacts and identifying where the intervention is or is not working.

Consider, as an example, the case of an intervention aimed at transferring more managerial authority to the school level. In this case, the intervention is intended to improve such elements of service delivery as teaching quality and teacher attendance, with the ultimate goal of improving student outcomes. Better measurement and evaluation are necessary all along this envisioned results chain. Of all of these elements, several indicators of student outcomes—enrollment, persistence, and completion—are probably measured most regularly and accurately, but even in these cases we often must rely on household surveys to supplement the administrative data. Student learning is typically poorly measured, and so improving

measurement of learning outcomes is a major thrust of the education-sector benchmarking exercise that the World Bank has launched.⁴

The other elements in the results chain—the governance intervention and the quality of service delivery—are generally not measured on a regular basis. With this initiative, our objective is to help increase the frequency and the consistency of measurement of those indicators. In the case of school-based management, for example, this would mean collecting data and doing qualitative assessments to determine to what extent and in what areas decision-making authority had been devolved to the school level, the extent of actual participation, the extent of teacher presence and teaching quality, and the extent of improvements in student learning outcomes.

All of this applies equally in the health sector. Consider, for example, the case of an intervention directed at paying provider organizations according to their performance and giving them autonomy to manage inputs. As with education, we have better measures of final outcomes than of governance and quality of service delivery. For example, governance measurement might include the extent to which the process of contracting of health services was competitive and free of corruption. Measures of service delivery quality could include the quality of advice provided by medical providers, perhaps through direct observation of interactions between providers and patients.

⁴ See www.worldbank.org/education/saber.

Better Measurement Can Improve Project Monitoring

Finally, better indicators for measuring service delivery can improve the monitoring of development projects and, ultimately, project quality. There are two main reasons why this is important. First, clients and the donor community are increasingly demanding tools to improve the results of development projects. And second, the design of projects is becoming more results-focused, with a growing number of projects linking disbursements to project outcomes.

Development partners increasingly recognize that improving project monitoring requires better measurement. Tracking inputs and outputs is not sufficient for improving human development outcomes. Instead, better measurement of service delivery and the policies that affect service quality are necessary to look inside the “black box” of what actually happens at the point of delivery. For example, stock-out rates for pharmaceuticals can indicate whether essential medicines are reaching health clinics. Such indicators can be incorporated into project monitoring arrangements for investment projects and can also be used as targets in adjustment operations.

The growing popularity of results-based projects also requires careful attention to indicators.⁵ These operations link disbursements to project outputs and outcomes, instead of inputs. For example, in education, a recent loan in Pakistan links disbursements to the extent of merit-based recruitment of teachers. In health, results-based projects link payments to coverage and health status outcomes. These examples require monitoring of intermediate policy indicators

⁵ For more on results-based financing approaches in health, see: www.rbfhealth.org.

to ensure that the project is on-track, as well as of the outcome indicators linked to disbursements.

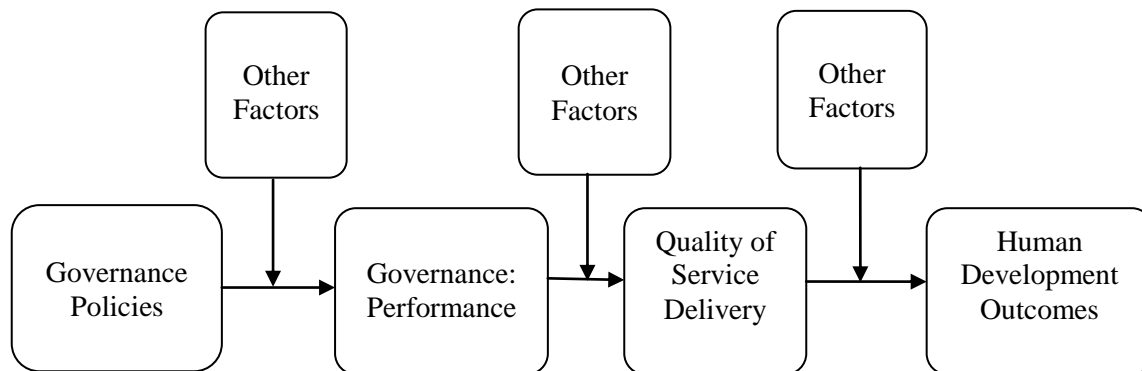
Finally, the use of governance and service delivery indicators for project monitoring can have the additional benefit of building the measurement capacity of countries. The inclusion of these indicators into results frameworks means that countries will have to collect this data over time. This provides a window of opportunity for incorporating governance indicators into administrative data and on-going surveys that can exist beyond the life of the project.

A Framework for Measuring Governance and Quality of Service Delivery

While there is great interest in defining and measuring the role that governance and quality of service delivery play in human development outcomes, there is limited consensus on how this complex relationship works in practice. Ultimately service delivery is only part of the story. Health status is influenced by outside factors, such as the quality of the water supply and infrastructure, while educational outcomes are influenced in part by family characteristics such as the level of parents' education. Behaviors of service users (e.g., adherence to medical treatment or time spent on homework assignments) are also critical in determining final outcomes. Similarly, governance is only one factor influencing the quality of service delivery. For example, the quality of medical care or teaching is influenced by factors as varied as the availability of technologies and the overall functioning of the labor market; for the most part, we would consider these factors to be independent of governance arrangements, at least as defined in this paper.

The discussion in this paper acknowledges this context and focuses on two interrelated factors driving outcomes: the quality of governance and the quality of service delivery. As shown in Figure 1, governance can influence the quality of service delivery, which in turn affects human development outcomes. Exogenous factors and other inputs contribute along the way.

Figure 1: Governance, Service Delivery and Outcomes



In this context, governance has two dimensions that can be measured. First, there is the set of *policies*, or rules of the game, that influence the ways that service providers function. These are the incentive and accountability arrangements. For example, human resource policies in health establish the rules and regulations for hiring, firing, and rewarding doctors; the same is true in the case of teachers in education.

Second, governance can be described in terms of *performance*—the ways in which policies actually influence the behaviors of providers. Governance performance is measured by

the extent to which the governance policies work in practice.⁶ In the personnel example, whether doctors or teachers show up regularly for work (e.g. attendance rates) are a measure of governance performance.

Governance policies and performance in turn influence the *quality of service delivery*. This refers to the quality at the actual point of contact between provider (the teacher, doctor, or other medical provider) and client (the student or patient). Measurement of quality is often elusive; however, recent research has made progress in measuring dimension of quality such as the quality of medical advice and the classroom time use of teachers. Because of the difficulty of measuring quality, indicators of quantity are sometimes used as proxies. For example, in education, the amount of time teachers spend in classrooms may be used as an indicator of quality, if the actual quality of teaching cannot be measured.

This distinction between *policies*, *performance*, and *quality of service delivery* can help in thinking through the role of governance along the service delivery chain. For example, in health care, governance policies include personnel rules that regulate whether a local government can hire and fire workers, or pharmaceutical management and procurement policies that determine the way in which drugs reach facilities. Whether or not these policies work in practice can be

⁶ It is important to note that governance performance indicators aim to go beyond simple “in practice” application of governance policies. Instead they capture evidence on the extent to which policies are followed. For example, a governance policy could require advertising of posts for recruitment of teachers. Whether posts are advertised is an “in practice” indicator, but it does not provide a great indicator of the performance of a recruitment and selection process in ensuring meritocratic hiring decisions. Instead, it captures whether the “in law” requirements are being followed. A better indicator of the “performance” of such a selection process would be the average number of qualified applicants per advertised vacancy. This is because an advertising requirement may be formally followed, but its purpose may be compromised if, for instance, potential applicants believe that the selection process is rigged, and that an already identified candidate is all but assured of winning the competition.

measured by performance measures such as absenteeism rates for personnel, or the availability of drugs. We can then ask whether the increased availability of staff and drugs have improved the quality of service delivery, which could be measured through the number of health consultations following acceptable standards, or the number of patients receiving prescribed medications. Similarly in education a policy that provides bonuses for teachers who work in remote rural areas (*governance policy*) may influence the presence or absence of teachers in rural areas (*governance performance*) as well as the amount of time teachers spend teaching in the classroom (an indicator of *service quality*). These measures of service quality in turn influence outcomes.

This paper uses this distinction between governance policies, governance performance, and quality of service delivery as a framework for discussing indicators for measuring service delivery in health and education. The distinction raises a number of conceptual and measurement challenges, including the difficulty of defining the scope of governance policies that are relevant for influencing provider behavior and the complexity of distinguishing between measures of governance performance and the quality of service delivery (Box 1).

Box 1: Governance Performance vs. Quality of Service Delivery

While there is considerable discussion among experts about whether to measure governance performance or the quality of services, in our view, as indicated in the discussion above, good measurement of both of these elements – as well as of education and health outcomes – is crucial to improving service delivery. Two examples make it clear that we cannot choose just one or the other:

Outcome	<i>Is driven in part by:</i> Service Quality/Performance	<i>Which in turn depends on:</i> Governance Policy
Health status of patients	Quality of medical advice from providers	Regulatory environment for medical sector
Student cognitive achievement	Classroom attendance rate of teachers and time-on-task	Extent & quality of school-based management

In the first case, the subsequent health status of those who seek medical care from providers depends in part on the quality of advice offered by those providers. It is important to have some direct measurements of that quality, in order to know whether health-care spending is able to lead to better health status. Research from several countries has revealed great shortcomings in the quality of care, so much so that in a significant share of cases, following the doctor’s advice would actually harm the patient. But once we have identified shortcomings, in order to remedy them it will be important to track down their source. This means measuring the quality of upstream governance that directly affects the quality of care. For example, one might ask why the regulatory environment is such that large numbers of unqualified doctors are able to operate with impunity.

In the second example, the evidence suggests that a teacher’s skills and motivation are the most important school-based factor determining students’ learning. But these factors have until recently been poorly measured: education officials and researchers have had to make do with measures like the education level and experience of teachers, which have been shown to be poor proxies for classroom effectiveness. It is important to get inside the black box of teacher factors by measuring more directly the quality of teaching actually provided in the classroom. As with health, it is also necessary to measure the quality of upstream governance factors that affect teacher behavior – for example, the extent and effectiveness of school-based management, which is hypothesized to improve teacher effort and performance.

Another important consideration is the need to specify the level and unit of analysis.

Governance can be measured at a system-wide level, but also at the local provider level. For example, in some systems, individual schools or hospitals may have their own human resource policies. Thus, a critical aspect of assessing governance systems is developing a clear understanding of the institutional arrangements, including roles, responsibilities, and the authority of key actors. This is particularly important in decentralized systems, where a

mismatch in authority between levels of government can lead to poor governance outcomes—for example if local governments are responsible for financing services but have no authority for holding providers accountable.⁷

Measuring Governance: Policy and Performance Indicators

Governance systems are multi-dimensional, and the rules of the game influencing the behavior of providers are often complex. To focus our work, we are proposing to concentrate within the governance arena on five dimensions of the service delivery process: *human resources, financing systems, critical inputs, information, and provider entry*. Together, these dimensions can help describe a governance system through its policies and performance.

Measurement, in this context, implies both assessing the rules of the game (that is, the policies—both formal/*de jure* and informal/*de facto*) in each of the five dimensions and capturing the effects of those rules on actual performance. As noted in Savedoff (2009), not all of the governance policies will have scores that are easily rankable, since we cannot be sure *a priori* which direction (that is, more or less of the variable) will be associated with better service delivery and outcomes. However, all are hypothesized to be related to the performance of governance in some model of service delivery.

⁷ Refer to Fiszbein (2001) for a framework for institutional analysis.

Indicators of governance performance, on the other hand, should capture the behavior change brought about by the governance policy. These are generally indicators for which there is a widely shared sense of directionality—that is, for which research has established whether more or less of the variable is desirable.

The following sections discuss a potential “long list” of indicators in each of the five subareas of service delivery, covering both indicators of policy and performance. The goal of this discussion is to provide a set of indicators that teams can draw from to incorporate into surveys and monitoring frameworks.

Human Resources

Human resources is a particularly important area for measuring the quality of governance. Health and education are labor-intensive services that involve many hard-to-monitor “micro” transactions between providers and users. As a result, the rules of the game that define recruiting, hiring, compensating, assessing, rewarding, placing, and firing public health and education workers, as well as the incentives that affect their behaviors, constitute a critical dimension in our framework. In addition, because provider compensation claims the largest share of the recurrent budgets in education, and a large share in health as well, human-resource policies have major implications for the efficiency of use of budgetary resources.

On the governance *policy* side, a range of policies and practices affect the quality of human resources in service delivery. These include policies governing recruitment, retention, and assignment; monitoring and evaluation of performance; employment status and job security;

salary structure and other benefits; and workload, duties and autonomy. Potential indicators include, for example, those measuring the degree of meritocracy in civil-service hiring procedures, the existence and use of performance evaluation procedures, and the share of pay of typical provider based on performance.

The World Bank has recently launched a major effort to carry out comparable assessments of the quality of teacher policies across many countries. This effort, known as SABER-Teachers, is part of the Bank's broader System Assessment and Benchmarking for Education Reform, which is assembling data on the quality of policies in each major domain of the education sector. Until now, there has not been any source for systematic, comprehensive data on teacher policies around the world.

The SABER-Teachers program tries to fill this gap by collecting information on 10 core teacher policy areas in education systems around the world, which correspond to indicators of governance policies in our framework.⁸ It also assesses how well those policies rate from the perspective of achieving eight teacher policy goals that are shared by education policymakers and beneficiaries in most countries because they contribute to learning and are actionable. These goals include, among others, attracting the best candidates into the teaching profession, setting clear expectations for teachers, monitoring teachers and learning, and motivating teachers to perform. SABER-Teachers classifies education systems as being more or less advanced in progressing toward each of these goals.

⁸ For details, see <http://go.worldbank.org/MU6QMF8340>.

This teacher policy assessment tool has already been applied or is now being applied in 30 countries across all the developing regions, and its application suggests how these governance policy indicators can be used. Once data are collected for a country, the Bank produces a teacher policy report that analyzes the quality of its teacher policies. For each goal, in addition to assessing the country's progress, the report provides information about successful examples and approaches from other countries. Country reports have so far been produced for over a dozen countries, primarily in East Asia and the Middle East and North Africa. They are being used by the Bank and governments to identify opportunities to improve the effectiveness of their policy framework. For example, a discussion of SABER-Teachers in India with the leadership of over 20 state Secretariats of Education forms a foundation for a new secondary education project supported by the World Bank.

What about the government *performance* side? Evaluating the effects of these governance policies will typically involve measuring abuse and poor work effort on the part of health and education workers, as well as assessing the ability of provider organizations to attract and retain qualified staff. Potential indicators of governance performance in this dimension include corruption in the allocation of teaching posts, the share of “ghost” providers (those who are on the central administrative records but not on the rosters of facilities), and the level of provider skills (actual measured skills, not qualifications on paper).

One increasingly commonly used measure of government performance in human resources is the level of absence of teachers, doctors, and other medical personnel. High levels of provider absence are both an indicator of general shortcomings in accountability within the

education or health system (World Bank 2003) and, it appears, also a direct cause of poorer outcomes (Miller, Murnane and Willett 2007; Duflo, Hanna and Ryan 2008). But careful study of the causes and effects of absence depends on accurate indicators of a system's performance on this indicator—that is, the actual absence rate. Only in recent years have surveys focused on carefully gathering reliable absence data through direct verification of attendance during surprise visits to schools and clinics, rather than relying on questionable administrative records of attendance.⁹

Academic researchers and World Bank staff developed and pioneered this methodology in a series of studies in the health and education sectors over the past decade, proving the feasibility and usefulness of this indicator. For example, a 2004 World Bank study based on visits to a representative national sample of health centers in Bangladesh showed how important direct measurement of attendance is, when it found that 74 percent of doctors posted to the most remote rural clinics were absent at any given time (Chaudhury and Hammer 2004). Another study measured absence in representative samples of primary schools and primary health care clinics in six countries, finding absence rates that averaged 19 percent among teachers and 35 percent among medical workers, with far higher rates in some Indian states (Kremer et al. 2005; Chaudhury et al. 2006; Muralidharan et al. 2011). Subsequent studies by Bank operational units applied this methodology to yield reliable provider-absence estimates and inform the policy dialogue in Ecuador (Rogers et al. 2004), Mongolia (World Bank 2006), and Lao PDR

⁹ Administrative records can be easily manipulated and therefore do not serve as a reliable source for absence data.

(Benveniste, Marshall and Santibañez 2007; World Bank 2008). Provider absence rates also constitute one of the core indicators adopted in the multi-country governance measurement effort being launched by the African Economic Research Consortium with World Bank assistance (Bold et al. 2010), and they have been included in recent proposals for governance indicators in health (Lewis and Pettersson 2009; Savedoff 2009).

Financing and Resource Management

The second essential element of governance in service delivery is *financing and resource management*. This element includes the rules that define budgets and govern the transfer of resources across agencies and levels in the service delivery process; the extent to which the use of resources is monitored and reported; and the influence that the rules have on whether resources are available and are used at the point of service delivery.

As a start, governance indicators for financing and resource management can build on the PEFA (Public Expenditure and Financial Accountability) framework. PEFA was developed by a donor-financed secretariat at the World Bank, which built on the Heavily Indebted Poor Countries (HIPC) initiative's efforts to develop indicators to monitor government expenditures and progress against HIPC commitments in countries' Poverty Reduction Strategy Papers (PRSP). Beginning in early 2002, the formal PEFA framework was developed, piloted in 24 countries through desk exercises, subjected to consultations, and then revised.

PEFA surveys, which are now regularly conducted in about 100 countries, summarize public finance performance using a set of 28 composite indicators. These indicators rank

countries on various aspects of public financial management, including budget credibility, transparency, and performance the budget cycle. The PEFA indicators include a combination of policy and performance indicators. For example, they incorporate *de facto* elements, such as the existence of a clear budget calendar, as well as *de jure* elements, such as whether the legislature followed the budget calendar in the year preceding the survey. Like the Bank's CPIA index, PEFA indicators are scored on a four-point scale (Lewis and Pettersson 2009; PEFA Secretariat 2005).

The PEFA framework looks only at indicators of financing policy and performance at the level of the overall government, not individually at sectors. At least two recent efforts have applied this approach at the sectoral level, for the health sectors in the Philippines and in Mozambique (Lawson et al. 2008; Philippines Department of Health).

The Philippines study, on which the Government of the Philippines and the World Bank collaborated, defined a set of Agency Benchmark Indicators (ABI) that assess the public financial management performance of the Department of Health (DOH). The exercise developed indicators of financial management practices, policies, procedures, and systems that can be applied at an organizational level and can be tracked over time. The Department of Health was the pilot agency, but the goal is to apply the instrument in other sectors as well. The indicators cover six areas: (i) budget preparation; (ii) budget execution and reporting; (iii) organizational management and accountability for performance; (iv) DOH-specific indicators; (v) external factors influencing DOH operations; and (vi) budget credibility, execution, and efficiency. These indicators include measures of both policies and performance. In the area of budget preparation,

for example, measures include whether DOH has budget preparation guidelines and are disseminated on time (policy), but also whether funds are released on time and how the capital budget is allocated (performance).

The Philippine ABI analysis provided policy-relevant findings that can be used to improve service delivery. For example, in the area of budget preparation, the assessment found that the process is constrained by frequent changes to the budget calendar and by the limited time that individual units are given to prepare their budgets. In the area of financial reporting, the analysis found that while the DOH publishes public reports about its programs, it makes available little of the information necessary to assess value for money. This effort shows the usefulness of an agency-specific tool for assessing financial performance. The ABI could be modified for use in other countries.

This type of analysis of financial management performance at the *organizational* level could be usefully complemented by tools for analysis of policies and performance at the *sectoral* level. In the health sector, for example, resources flow to local health facilities from central and sub-national governments as well as, in many cases, a health insurance fund. Each of these may have its own rules, institutional culture, and level of efficiency. Comprehensive sector analyses such as public expenditure reviews (PERs) and public expenditure tracking surveys (PETS) can complement organizational analyses to provide a more comprehensive assessment of the quality of financing of service delivery. PERs and PETS are also useful tools for measuring the performance dimension of financing and resource management (Koziol and Tolmie 2010). PETS measure leakage by tracking resources from the central budget to the facility level.

Procurement and Management of Critical Inputs¹⁰

The rules regarding what critical education and health inputs are procured, how and by whom they are procured, and what processes regulate availability, quality, and cost constitute another important dimension of governance systems. While health and education services require many inputs, textbooks and especially pharmaceuticals are often considered critical for quality of service. Here we discuss indicators related to pharmaceuticals, an area in which there has been particular interest both at the international and national level in many developing countries.

The World Health Organization has developed a methodology for assessing the transparency of pharmaceutical policies that can serve as a useful example of a multi-dimensional, expert-opinion-based system of measurement of governance policies. It covers 8 dimensions: medicines registration, licensing of pharmaceutical establishments, inspections of those establishments, control of medicine promotion, control of clinical trials, selection of medicines, procurement and distribution (World Health Organization 2009).

In each of these dimensions, the WHO defined a set of indicators to assess the transparency of the policies that govern that specific area. For example, for the area of medicines registration, the assessment considers such questions as:

- Is there an up-to-date list of all registered pharmaceutical products available in the country?
- Does it provide a minimum level of information?

¹⁰ Health indicators draw from WHO, 2008.

- Is there a standard application form publicly available for submission of applications for registration of medicinal products?
- Is there a functioning formal committee involved in the assessment of the applications for registration of pharmaceutical products?
- Are there clear written criteria for selecting the members of the committee?

In the area of procurement, the WHO considers whether there is written guidance for procurement office staff on the type of procurement method to be used for different types of products, or if there is a formal appeals process for applicants who have their bids rejected. Similarly, it asks whether there is a tender committee and, if so, whether the key functions of the procurement office and those of the tender committee are clearly separated. Scores are defined for each indicator and aggregated for each dimension.

The explicit assumption of the WHO approach is that these governance policies (focused on transparency) define the system's degree of vulnerability to corruption. The methodology has been applied to 25 developing countries. The findings of those assessments are summarized in a recent report, which finds that the dimensions of greatest vulnerability are inspection, promotion, and selection (Kohler and Bahdadi-Sabeti 2011). Perhaps surprisingly, it identifies procurement and distribution as the areas with the least vulnerability to corruption.

There have also been efforts to develop performance indicators related to pharmaceuticals. One area of performance that has received attention is the share of pharmaceutical sales that consist of counterfeit drugs. This indicator reflects the quality of the drug supply and the effectiveness of quality assurance policies such as inspections.

A USAID project called "Promoting the Quality of Medicines" collects and test samples of drugs at various outlets in many countries, to check for counterfeit drugs, and then publishes

the findings in a database¹¹. This database provides detailed information, including the stated name of manufacturer and lot number, dates of collection and testing, types of tests performed, and names of the testing facilities. The Pharmaceutical Security Institute also collects information on incident trends, although its database is less detailed.

Stock-out rates (absence) of essential drugs in health facilities are a much-cited indicator of the availability of essential drugs in health facilities. When measured over enough time, this indicator can help assessing whether absence of drugs reflects a supply or a demand issue. Not all factors influencing stock-out rates are related to governance, but given the often serious problems of abuse that plague the distribution and use of pharmaceuticals, the policies discussed above are expected to influence stock-out rates.

The World Health Organization (WHO), in collaboration with the global NGO Health Action International (HAI), compiles data on the prices and availability of essential medicines in many countries. HAI collects data on price and availability on a list of essential medicines (defined by WHO) through surveys carried out in six different regions in each country. In larger countries, it collects data at the state level. The sample consists of medicine outlets in the public, private, and other (non-profit) sectors in the country. The survey data are available to the public through a database maintained by HAI.¹² The database contains information on drug pricing and availability for 49 countries with 15 from Sub-Saharan Africa, 7 from Latin America and

¹¹ See <http://www.usp.org/worldwide/medQualityDatabase/>

¹² <http://www.haiweb.org/MedPriceDatabase/>

Caribbean, 12 from Middle East and North Africa, 6 from Europe and Central Asia, 3 from South Asia, and 6 from East Asia and Pacific regions.

The focus on stock-out rates can also be seen in the active campaign by non-governmental organization in several African countries.¹³ Using SMS technology, these NGOs collect disaggregated stock-out indicators, which they then present and disseminate in creative graphical form.

Information

Decisions and behaviors of the various actors along the service delivery chain are based on the information that those actors have on inputs, outputs, and outcomes. The rules regarding which information is collected and made available, as well as the availability, reliability, and timeliness of that information, can influence governance performance. Other important factors include the extent to which the views of current and potential users are captured and considered in the design and implementation of services (for example through grievance processes or participation mechanisms such as school committees or) and the level of access that beneficiaries have to information about services (for example through publication of test scores of individual schools). Governance indicators in the area of information can therefore look at the quantity and quality of information available; the amount of access to information for stakeholders; and the availability of feedback mechanisms.

¹³ See <http://www.stockouts.org>.

On the policy side, the existence of a legal framework providing for transparency is thought to strengthen accountability by disseminating to citizens information about their rights, and service standards and performance. So far 80 countries have adopted access-to-information Laws, and approximately 50 have access-to-information rights included in their constitutions.¹⁴ This raises the question of whether and how the existence of such legislation affects service delivery at the sectoral level, and what indicators of information access—both of policy and performance—can and should be monitored.

The World Bank and the NGO Global Integrity have developed a set of indicators for assessing transparency and access to information at the sector level in health and education.¹⁵ These indicators were tested in FYR Macedonia and will be collected in Ukraine and Kenya in 2011. The survey uses an expert assessment methodology based on interviews with teachers and doctors and parents and users of health services. The indicators are clustered around four dimensions of information access to health and education service delivery:

- *Existence and Usability of Information* assesses the availability, accessibility, and usability of information on health and education services, including whether information on quality of performance is accessible in user-friendly format and whether information is standardized in a way that makes it comparable across providers.
- *Redress Mechanisms* looks at the availability and accessibility of complaints-handling mechanisms and institutions, as well as the time taken to lodge a complaint.
- *Availability of Fiscal/Budget Information* looks at the availability of basic fiscal and budgetary information that would theoretically allow local citizens (often through

¹⁴ See <http://right2info.org/constitutional-protections-of-the-right-to>

¹⁵ Details are available at <http://commons.globalintegrity.org/2011/01/new-data-information-access-in-health.html>.

intermediaries such as CSOs and media) to monitor service delivery resource flows and the allocation of funds.

- *Citizen Participation in Local Decision-Making* examines the existence and effectiveness of formal consultative mechanisms, as well as other informal mechanisms that could theoretically convey citizens' concerns effectively to policymakers.

The indicators look both at policies (including the existence of formal provisions for making information available and redress mechanisms) and at performance (whether people access information and file complaints in practice).

The indicators are useful for mapping and identifying what information related to rights, institutions, and mechanisms is available and how effective the system is at providing health- and education-related information to the beneficiaries of services. In the case of FYR Macedonia, the data show a considerable gap between information access laws affecting education and health services and their implementation and enforcement. While a substantial portion of the relevant legal framework is in place, significant work remains to be done to implement the legislation.

For example, although the public has the legal right in Macedonia to access school budgets, this right is regulated by a budget law rather than a law on education. As a consequence, while budgets are made available to the public, they do not include full details such as an itemized list of budget allocations. The data identified a similar disconnect on complaint mechanisms. While a Law on Ombudsman provides a redress channel, it is not widely used for problems with services, and people are more likely to complain informally to family and through their local networks.

Provider Entry

The ability of new providers to enter the market for health and education services can be an important determinant of outcomes in the sectors. Private non-profit and for-profit providers, as well as new public-sector entrants (such as charter schools), can affect the quality of service delivery in two ways: by providing care directly, and by exerting competitive pressures and spurring behavioral change in public providers. These effects can be either positive or negative and reflect a trade-off between quality control and competition. On the one hand, requirements for entry can ensure that only qualified service providers practice; on the other hand, requirements that are poorly designed or applied may prevent qualified schools or health clinics from practice.

Few studies have assessed the overall conditions for provider entry into the delivery of health and education services.¹⁶ General indicators of ease of starting and running a business, such as those taken from the World Bank *Doing Business* survey, may capture some of the general business climate variables that could affect whether it is feasible and potentially profitable to run a (for-profit) school. But they do not provide a solid basis to understand either governance policies or performance in health and education with the necessary granularity.

¹⁶ Harding and Preker (2003) discuss the core elements of such an assessment for health services. An example of the conditions for private-sector involvement in health delivery can be found among the indicators for USAID's PSP-One program in health (Private Sector Partnerships-One Project 2005). Similarly, African Private Schools Investment Index (School Ventures and Economist Intelligence Unit 2008) provides an assessment of the ease of starting up a school.

A very recent effort to generate and collect indicators of how governments engage with the private sector to improve health in Africa (International Finance Corporation (IFC) 2011) provides a number of indicators that would be useful for our proposed measurement framework. Their assessment is broader in purpose and covers a range of domains (including the existence and nature of a policy framework and dialogue on private sector issues).

Two of the IFC's domains are of particular relevance for our purposes. The first assesses the inclusion of the private sector in the national health information system—one of the most basic indicators of whether private providers are truly incorporated into a health system. They find that in 12 out of 45 countries, private providers are not mandated by law or regulation to provide health authorities with basic information such as births or deaths or health service utilization data. Out of the 33 countries that do have the mandate, the information reaches the ministry in only three. They go on to argue that in some cases requirements are onerous, which may explain the low rate of compliance.¹⁷ These findings imply that both the lack of information policies and the inadequacy of those that do exist is weakening governance performance, as measured by the availability of basic information on private sector provision.

The second relevant domain in the IFC framework examines regulations governing the quality of private provision. This domain assesses whether a country sets requirements for registering a private clinic, has procedures for inspecting those clinics, and has standardized rules for operating a private clinic. It also assesses the quality of those rules and procedures, for

¹⁷ In the case of Burundi, for example, the requirement involves completing 27 forms per month.

example by determining whether the registry is updated regularly and whether the rules for operation are appropriate and reasonable. The assessment indicates that overly complex frameworks that are contradictory or cannot be implemented as intended create uncertainty and opportunities for arbitrary enforcement. The report argues that even private providers complain about the lack of oversight, which allows low-quality providers to continue to operate.

While it relies on surveys of experts in each country for information on both governance policies and performance, the IFC study tries to ensure comparability by providing a detailed and clear set of coding principles for all indicators.

Measuring Service Quality: A Parallel and Related Agenda

At the same time, governments and donors will need indicators of the quality of service delivery. Conceptually, what we want is a set of indicators that let us know how effective the delivery of services is—in effect, how much the services can be expected to improve the health or education of the client. These quality measures can differ from the governance performance measures described in the previous section, because other non-governance factors also affect quality (as represented in Figure 1). The line separating these indicators from the governance performance indicators is a blurry one, but it is useful to try to distinguish the two. For example, an important determinant of the quality of schooling could be whether or not there is a roof over the students' heads. While the availability of school buildings ultimately depends on governance, it would be a stretch to consider this one of the short- or medium-term “governance

performance” variables. Our *quality* measure could implicitly take into account the effects of having or not having a roof, whereas the governance *performance* indicators would not.

Service-quality indicators that have been used so far are of two types—objective and subjective.

Objective Indicators

Measuring the quality of service delivery is challenging. Even measures of governance performance, though only recently developed, have received more attention than quality measures. What do we know about how to measure service quality objectively?

In health, a set of recent studies has tried to measure the *quality of health care delivered* by observing doctor behavior (summarized in Das, Hammer, and Leonard 2008). These studies focus on sets of symptoms for which the appropriate treatment is clearly known, so that they can assess doctors’ performance against an objective standard. This allows comparison of quality across different sectors and environments—public vs. private clinics, for example, or rural vs. urban. Governance factors affect quality of care, but so could other dimensions like the clinic location and environment.

An important contribution of these studies is their ability to distinguish the twin determinants of quality of care by providers: skill and effort. Specifically, the studies use medical “vignettes” (or hypothetical cases) to measure the doctors’ knowledge of appropriate treatments, but then directly observe those doctors’ dealings with patients to see whether the doctors apply this knowledge. They take the gap between knowledge and practice to indicate

weaknesses in providers' effort, which is a crucial dimension of quality of care. *Effort* can also be considered an indicator of governance performance, as it is influenced by incentive arrangements (such as fees for service or performance bonuses), although other factors like intrinsic motivation also affect it.

In education, there are corresponding measures of the quality of teaching, gathered through *classroom observation* studies. One such indicator is *time on task*—the total amount of time that teachers are actively engaged in teaching and students are actively learning (for example, Abadzi 2006). Most observers will agree that schooling is likely to be more effective when teachers are able to devote less time to maintaining discipline or carrying out administrative tasks and more time to teaching. However, a concern with time-on-task studies is the amount of variation by enumerator. It is difficult for researchers to agree on how to categorize how teachers spend their time in the classroom. Beyond time on task, if we are willing to make assumptions about the most effective pedagogy (for example, by prioritizing “active learning”), then the time-on-task metrics could be made more detailed by measuring the amount of time spent on good pedagogy. These metrics could be used to construct more direct *assessments of the quality of teaching* as indicators of quality of education services.

Subjective Quality Measures

A second type of indicators is measures of satisfaction of the clients or recipients of health and education services. These subjective measures can proxy for measuring actual quality of services and can also measure the responsiveness of services to the needs and preferences of clients. In the case of health services, we can survey patients about their experience, for example

through exit interviews at clinics. In the case of education, one could survey the students, but it also makes sense to survey parents and other stakeholders (including employers) for their opinions of how well schools are functioning.

Such indicators need to be interpreted with caution, for several reasons. First, to take the education example, parents may have different objectives for their children's education than society as a whole does. Especially given that public provision of services is often justified by the presence of externalities, it would be theoretically inconsistent to argue that the user's subjective measures capture all desired outcomes. For example, parents may be happy with a teacher who reinforces traditional gender roles even as the government tries to open up opportunities for girls (or vice versa). Second, even if parents and society share the same goal—such as more rapid student learning—parents may have been conditioned to have low expectations, leading them to report high satisfaction despite poor schooling. A third reason for caution is information asymmetries; these may be especially important in health services where providers have technical information and expertise that patients lack. As a result, patients' perceptions of whether they are receiving quality care may be based on factors not relevant to technical quality. For example, they may believe they are receiving quality care if the clinic is clean and the doctor is friendly, even if the doctor prescribes the wrong medications.

Despite these concerns, there is value in measuring users' opinions as a complement to the objective measurements. Ultimately, users should know at least what increases their utility better than the government or researchers will, as long as we take account of the possible biases noted above. Inconsistencies between the two types of indicators could be a useful warning flag

about possible measurement problems in the objective measures, or at least a prompt for deeper investigation.

One possible subjective indicator of service delivery quality would be reports of problems taken from *international surveys of individual respondents*. An example is the questions fielded by Gallup in a sample of countries in 2003-2006, which asked about the incidence of problems including poor teaching, overcrowded classrooms, school facilities in poor condition, lack of drugs, and disrespectful staff.¹⁸ Because they are included in general-purpose surveys, these questions will typically not yield detailed information, but may serve as a general barometer of quality of services. A second set of indicators would be the results of *household survey modules* on availability and quality of services. Household surveys could go into more depth than the polling-style surveys, and could produce more rigorously quantitative measures of performance.

Conclusion and Next Steps

In this paper, we have laid out a rationale and framework for constructing indicators of the quality of governance policies and performance in the health and education sectors, as well as the quality of service delivery itself in those sectors. We have also begun to fill in this framework by suggesting candidate governance indicators (both policy and performance) in five key areas—human resources, financing and management, information, provision of critical

¹⁸ The survey also included questions on provider absenteeism, which could be used under the “governance performance” indicators.

inputs, and provider entry—as well as a smaller number of possible indicators of service delivery quality.

To advance this agenda, the next step will be to begin gathering the data for these indicators more systematically. To do so effectively, it will be necessary to develop new instruments, and to make use of multiple types of instruments rather than trying to rely on one or a handful. Governance involves a complex set of processes that require different measurement approaches; moreover, much can be gained from cross-checking different sources. There are many promising instruments for collecting data: facility surveys such as Public Expenditures Tracking Surveys (Reinikka and Smith 2004), expert surveys on system variables, household surveys, secondary sources, administrative data, and qualitative studies. Describing these instruments is beyond the scope of this paper, but two recent surveys—Lindelov and Wagstaff (2008) and Amin and Chaudhury (2008)—offer excellent detailed discussions of instruments and methodologies appropriate for the health and education sectors, respectively.¹⁹

A second set of questions concerns who should carry out the data collection, and how often. Is it better to have the government collect data for these indicators, or to rely on an NGO or other outside actor? And should data collection and indicator production and dissemination be occasional one-off events, to raise awareness and set some benchmarks, or institutionalized elements of regular monitoring activities? While there are tradeoffs in each case, what is most

¹⁹ The edited volume by Amin, Das, and Goldstein (2008) reviews the instruments and lessons from their application to various cases. See also the working-paper version of this paper, which discusses the issue in more detail.

important is not coming to consensus on these details, but agreeing that it is time to begin the process of collecting governance and quality indicators in service delivery.

References

- Abadzi, Helen. 2006. *The Economics of Instructional Time: How Efficiently are Funds Converted into Learning Activities?* Washington, DC: World Bank.
- Amin, Samia, and Nazmul Chaudhury. 2008. An Introduction to Methodologies for Measuring Service Delivery in Education. In *Are You Being Served? New Tools for Measuring Service Delivery*, edited by S. Amin, J. Das and M. Goldstein. Washington, D.C.: World Bank.
- Amin, Samia, Jishnu Das, and Markus Goldstein, eds. 2008. *Are You Being Served? New Tools for Measuring Service Delivery*. Washington, D.C.: World Bank.
- Chaudhury, Nazmul, and Jeffrey S Hammer. 2004. Ghost Doctors: Absenteeism in Rural Bangladeshi Health Facilities. *World Bank Economic Review* 18 (3):423-41.
- Das, Jishnu, Jeffrey Hammer, and Kenneth Leonard. 2008. The Quality of Medical Advice in Low-Income Countries. *Journal of Economic Perspectives* 22 2:93-114.
- Harding, April, and Alexander S. Preker, eds. 2003. *Private Participation in Health Services*. Washington, DC: World Bank.
- International Finance Corporation. 2011. *Healthy Partnerships: How Government Can Engage the Private Sector to Improve Health in Africa*. Washington DC: World Bank Group.
- Kohler, Jillian Clare, and Guitelle Bahdadi-Sabeti. 2011. *The World Medicines Situation 2011: Good Governance for the Pharmaceutical Sector*. Geneva: World Health Organization.
- Koziol, Margaret, and Courtney Tolmie. 2010. *Using Public Expenditure Tracking Surveys to Monitor Projects and Small-Scale Programs: A Guidebook*. Washington, DC: World Bank.
- Kremer, Michael, Karthik Muralidharan, Nazmul Chaudhury, Jeffrey Hammer, and F. Halsey Rogers. 2005. Teacher absence in India: A snapshot. *Journal of the European Economic Association* 3 (2-3):658-67.
- Lawson, A., M. Umarji, T. Cammack, J. Guilherme, A. Nordine, A. Lala, and A. Makda. 2008. Assessment of Public Finance Management and Procurement Systems in the Mozambique Health Sector. Report to the Ministério da Saúde, the Ministério de Planificação e Desenvolvimento and the Ministério das Finanças and to the Health Sector Group of Co-operating Partners.
- Lewis, Maureen, and Gunilla Pettersson. 2009. *Governance in Health Care Delivery: Raising Performance*. Washington, DC: World Bank.

- Lindelow, Magnus, and Adam Wagstaff. 2008. Assessment of Health Facility Performance: An Introduction to Data and Measurement Issues. In *Are You Being Served? New Tools for Measuring Service Delivery*, edited by S. Amin, J. Das and M. Goldstein. Washington, D.C.: World Bank.
- PEFA Secretariat. 2005. Public Financial Management: Performance Measurement Framework. Washington, DC: World Bank. Public Expenditure and Financial Accountability Program.
- Philippines Department of Health. Financial Management Policies, Procedures, Practices and Systems: Agency Benchmarking Indicators Study.
- Private Sector Partnerships-One Project. 2005. Performance Monitoring Indicators for the Private Sector Program (PSP), with Explanatory Notes. Bethesda, MD: Abt Associates.
- Reinikka, Ritva, and Nathanael Smith. 2004. *Public expenditure tracking surveys in education*. Paris: International Institute for Educational Planning.
- Reinikka, Ritva, and Jakob Svensson. 2005. Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda. *Journal of the European Economic Association* 3 2-3:259-67.
- Savedoff, William D. 2009. Governance in the Health Sector: Authority, Information and Motivation. Portland: Social Insight.
- School Ventures and Economist Intelligence Unit. 2008. The African Private Schools Investment Index 2008. Alexandria, Virginia: School Ventures Inc.
- World Bank. 2003. *World Development Report 2004: Making services work for poor people*. Washington, DC: Oxford University Press for the World Bank.
- . 2008. Doing Business 2009: Comparing Regulation in 181 Economies.
- World Health Organization. 2009. Measuring transparency in the public pharmaceutical sector: Assessment instrument. Geneva: Departments of Essential Medicines and Pharmaceutical Policies & Ethics, Equity, Trade and Human Rights, WHO(<http://apps.who.int/medicinedocs/documents/s16732e/s16732e.pdf>).