



PART 5

**Government Analytics
Using External
Assessments**

CHAPTER 27

Government Analytics Using Household Surveys

Faisal Ali Baig, Zahid Hasnain, Turkan Mukhtarova, and
Daniel Rogger

SUMMARY

This chapter presents a guiding framework for using household survey microdata, readily available to most governments, to develop insights into the structure of the public sector workforce and the qualities of its compensation practices. National statistical authorities frequently collect household surveys with detailed information on labor force participation. These surveys are broadly consistent across time and are developed using globally standardized definitions and classification nomenclature. This offers governments unique insights into the public sector workforce that cannot be derived solely from administrative data sets, including the ability to juxtapose the demographics and skills composition of the public sector workforce to the private sector and assess the relative equity and competitiveness of public sector compensation practices. The chapter provides illustrations of the insights into public sector employment and wages that can be generated by this framework, using examples from the World Bank's Worldwide Bureaucracy Indicators (WWBI). Such insights can inform policy choices related to managing human resources in the public service.

ANALYTICS IN PRACTICE

- Governments routinely conduct household surveys in order to understand the populations they serve, target public policy, and inform policy debates. Many of these surveys record a dedicated and detailed set of variables on the labor market experiences of people in the country, including whether respondents work in the public or private sector. By collecting comparable data across the two sectors, household surveys provide a foundation for understanding the characteristics of public officials compared with their private sector counterparts—which is not possible through administrative data sources alone. Including precise and coherent indicators of the employment sector in household surveys thus enables their use in understanding the characteristics of public sector workers.

The authors' names are listed alphabetically. Faisal Ali Baig is a consultant, Zahid Hasnain is a lead governance specialist, and Turkan Mukhtarova is a consultant in the Governance Global Practice of the World Bank. Daniel Rogger is a senior economist in the World Bank's Development Impact Evaluation (DIME) Department.

- Given the unique nature of the public sector, time-series and international benchmarks and comparisons are critical for understanding the current state of government functioning. The precision and consistency with which household surveys are conducted across time open up the possibility of understanding the longitudinal dynamics of the public sector relative to the private sector. Similarly, ensuring such surveys collect data in line with globally standardized definitions and classification nomenclature enables comparisons across countries.
- Setting up household surveys in this way allows the government to juxtapose any feature of individuals in the public and private sectors that surveys have collected data on. Demographic variables allow for an analysis of how gender-related differences in pay vary across sectors, regions, organizations, and so on. Assessments of the skills composition of the public sector workforce with respect to the private sector identify in what areas the government is competing most intensively for skills with private sector actors and what that competition is doing to wages.
- Taking household surveys as the foundation of analytics around the equity and competitiveness of public sector compensation practices is a relatively low-cost approach to the analytics of the personnel determinants of state capacity. However, the usefulness of these surveys is underpinned by the representativeness of sampling and interviews, which ensure that the resulting data provide a robust estimation of labor force participants. This requires coordination between agencies of public employment and national statistical agencies.
- Detailed data from household surveys on the distribution and remuneration of public employees can help identify more nuanced, targeted, and politically feasible reforms that make explicit the difficult trade-offs in employment and compensation policies. Such an evidence-based approach is necessary because, historically, public sector employment reforms have often occurred in the context of economic crises, with an emphasis on blunt, short-term fixes that can have adverse impacts on long-term growth and welfare, and often create distortions and perverse incentives.

INTRODUCTION

The effective management of public sector employment and compensation is a vital activity of governments, with broad implications for fiscal sustainability, public sector productivity, and the competitiveness of the overall labor market. The wages of public sector employees consume a significant proportion of government expenditures. Across the world, government expenditures on employee compensation represent, on average, 30 percent of total expenditures (Hasnain et al. 2019). Spending on public sector salaries comes at the opportunity cost of spending on public sector programs.

At the same time, human resources in the public service are essential providers of government services and infrastructure, as well as ensuring the effectiveness of regulations (Arizti et al. 2020; Ingraham, Joyce, and Donahue 2003; Moynihan and Beazley 2016; Rasul and Rogger 2018). The size and nature of public sector wages affect the selection, retention, and motivation of public sector workers, which, in turn, impacts productivity, the amount and quality of government outputs, and public service provision (Finan, Olken, and Pande 2017).

These issues matter not only because they impact the quality of government functioning. The public sector is a large employer, accounting for, on average, 37 percent of global formal employment (Baig et al. 2021). Changes in government wages are likely to produce significant effects on the national labor market and the overall economy, including potentially crowding out recruits in the private sector (Behar and Mok 2013). In many lower-middle-income countries, especially those experiencing fragility, public sector employment is the core ingredient of the political settlement, and wage bill reforms have immediate and often severe implications for political stability, peace, and security (Gifford 2006).

There are thus several important questions about the public sector workforce that governments regularly need to address. What is the appropriate level of employment in the public sector as a whole and for essential workers like public administrators, teachers, and doctors, in particular? Does the public sector pay competitive wages compared to the private sector to attract talent while not crowding out private sector jobs? Does the public sector promote gender equality in employment, both in absolute terms and relative to the private sector? And are public sector pay and employment practices contributing to robust and dynamic labor markets at the national and regional levels?

Answering these questions requires high-quality microdata on public sector employment and compensation and comparable data for the private sector. Utilizing household surveys as a source of information on public employment offers certain advantages over administrative data. These data are routinely collected by national statistical organizations for informing broader policy goals and thus represent a cost-effective tool for government analysts. Household surveys provide a rich, consistent, and regularly updated set of variables for a variety of worker characteristics in the public and private sectors that enable robust, controlled comparisons between the two groups. Such surveys allow data to be drawn from the public and private sectors in a common manner. Thus, these data often represent a richer source of insights than are available from administrative data alone.

This chapter is targeted at government officials, development practitioners, and researchers who aim to gain a better understanding of the structure of the public sector labor market and its implications for the overall labor market. It begins by introducing the advantages of this survey-based framework and key areas for caution. It showcases the main features of the methodology and presents guidance for conducting analysis to delineate trends from these surveys. It goes on to provide guidance on how improvements in the design and conduct of labor force surveys allow for even more granular analysis. Finally, it illustrates the breadth of insights that can emerge from a study of public administration founded on household survey data. The approaches outlined here are a natural complement to those presented in chapter 10 of *The Government Analytics Handbook*, trading off the granularity of data with comparability with the private sector.

THE POWER OF USING HOUSEHOLD SURVEYS FOR GOVERNMENT ANALYTICS

Nationally representative *household surveys*, collected by national statistical authorities, are some of the most professionally conducted surveys in the world and are frequently supported or improved through consultations with multilateral organizations' data teams, which possess substantial experience in such exercises. By collecting data on a representative sample of the whole or some subset of the population, such surveys provide a window into the lives of those about whom data are collected. When surveys collect data on whether respondents work in the public or private sectors, they provide windows into life in those sectors.

This chapter will focus on employment-related variables in such surveys and thus on the *labor force modules* in household surveys. When these modules are the dominant concern of a survey, the survey is typically classified as a *labor force survey*. In much of the rest of the chapter, we will use these terms interchangeably. However, most of the principles of the discussion carry over to other elements of standard household surveys, such as consumption patterns drawn from consumption modules.

The Strengths of a Survey-Based Approach

Governments routinely rely on household survey sources to generate headline indicators of the health of labor markets and the overall economy. Insights emerging from these surveys are often used in the design of a wide array of economic and social policies. For example, information on the share of employed (and unemployed) individuals within the labor force frequently has direct consequences on the monetary and fiscal

policies of the government. Unemployment rates are often used as proxies for the vitality of the labor force and are used—in combination with inflation rates—in determining interest rates by central banks.

The quality of these data makes them an attractive foundation for government diagnostics of relative labor market characteristics and dynamics. Specifically, utilizing nationally representative labor force survey data to characterize public and private labor markets offers government and analysts five unique advantages over other data sources.

First, labor force surveys provide a rich, consistent, and regularly updated set of variables for a variety of worker characteristics in the labor market. Given the investments governments have made in methodological rigor, effective implementation, and quality assurance, these surveys are one of the richest available sources of information on population characteristics. Household surveys provide coherent descriptions of the composition of individuals within households, their demographics and qualifications, their consumption behaviors, and the nature and sector of their participation in the labor market, as well as detailed indicators on the industries and occupations they are engaged in and their salaries and other sources of compensation (including in-kind payments, government assistance programs, and social security benefits).

Second, labor force surveys undertake the same measurement approach across the public and private sector labor markets. This is a unique advantage of these surveys for measuring state capacity because these data are collected simultaneously for workers in both sectors from the same sample frame in a coherent manner. Administrative data sources (while being a potentially more accurate and detailed measure of employment and wages in the public sector) only include information on public sector employees and, at times, only the employees of particular ministries or organizations. It is extremely unlikely that any single administrative data set would not only cover workers employed across a diverse set of economic activities (from agriculture to mining, manufacturing, and the services sector) but also include information on both public and private sector labor force participants.¹ Even if such a data set does exist, its data will rarely be consistent with administrative data from other countries, complicating international comparisons.

Third, the granular nature of the underlying data ensures that labor market models are based on representative data sourced from across the economy. These surveys often sample thousands of employed individuals and are based on a meticulously designed sampling frame based on national census data, allowing for a close and accurate approximation of local labor markets. This reduces the assumptions on which analysis is based—the data are allowed to speak for themselves—and allows for decomposition by the characteristics of workers where sample sizes are sufficiently large.

Fourth, household surveys may represent a more complete view of the public sector workforce than even administrative data sets. Public sector administrative data are often too restrictive in defining who is included in their measurement. For example, contract workers have become an essential part of the public sector workforce, working alongside permanent staff in the promulgation of regulations and the delivery of social services. In many developing countries, they also represent a sizeable proportion of the public sector education and health care workforce. However, they are not counted as formal public employees in many administrative systems; that is often the reason for their contractual status. Given that contract workers are often exempt from budgetary limits on spending, their recruitment allows ministries to circumnavigate budgetary conditionalities against overspending on personnel.

This impairs the ability of governments to assess the true size of the public sector workforce. Further, given that these data sets are often unique to individual public sector organizations, the determination of who does or does not constitute a public sector worker may differ across organizations. Both of these factors would bias any estimate of the size of the public workforce, weakening the ability of governments to track wage bill spending. Survey data, on the other hand, are not limited by this distinction. Given that they are collected directly from individuals, who can elaborate on their sector employment, surveys can allow for a better determination of the size and structure of the public sector workforce.

Finally, household survey data are typically collected with research and diagnostic objectives in mind. Administrative data sets are collected for a variety of nonstatistical reasons, such as human resource management, program administration, or other regulatory or enforcement purposes. Therefore, administrative data in their “raw” form may not be suitable for statistical analysis.

Areas to Be Cautious in a Survey-Based Approach

It is important also to point to the caveats associated with the use of survey data. Labor force surveys can only generate insights into the nature and organization of public sector human resources with rich, complete, and consistent data collection on labor force participants, in general, and the public sector workforce, in particular. Given the self-reported nature of household surveys, respondents may not be able to fully articulate or comprehend nuances within their responses around the nature of their employment. While these surveys are designed and extensively piloted with particular care given to how questions may be interpreted, in order to ensure the quality of responses, there may be lingering imprecision within the variables that define individuals as working in specific sectors.

Utilizing a more broadly defined public sector identifier may make it easier for respondents to accurately answer relevant survey questions and allow for a more comprehensive comparative analysis between the public and private sectors. However, this may make the survey's definition of the public sector unfit for particular purposes. It is often difficult to differentiate between, for example, federal and state employees or those that are employed within specific ministries and those employed within state-owned enterprises.

The second area for caution is the respondents' representativeness of the underlying population of public and private sector workers. Labor force surveys, by definition, sample the working-age population. When designing the sampling frame, surveys often strive to ensure a representative sample in terms of age, gender, and racial and ethnic demographics. Some surveys aim to sample a representative share of employed and unemployed individuals and those not active in the labor force. However, they rarely if ever explicitly attempt to ensure a balanced sample of public and private sector workers. Analysts must thus assess whether the sampling approach might have biased data collection toward one or the other sector's employees or otherwise changed the nature of measurement in either sector.²

SETTING UP THE ANALYSIS

Capitalizing on Current Household Surveys

What features of a survey are necessary for it to be useful for government analytics? The answer to this question will be determined by the specific analysis intended. This chapter therefore follows the requirements of an analytical framework used by the World Bank to understand public and private sector labor markets in the development of the Worldwide Bureaucracy Indicators (WWBI). The WWBI is a unique cross-national data set on public sector employment and wages that utilizes global repositories of household survey data from 202 economies to present a globally consistent and analytically rigorous set of indicators across five categories: the demographics of the private and public sector workforces, public sector wage premiums, relative wages and pay compression ratios, gender pay gaps, and the public sector wage bill.³

WWBI indicators on public employment track key demographic characteristics, including the size of the public sector workforce (in absolute and relative numbers), the age of the workforce, and the distributions of employees across genders, industries, income quintiles, and academic qualifications. Compensation variables capture both the competitiveness of public sector wages (compared to the private sector) and wage differentials across industry or occupation, gender, education, and income quintile within the public and private sectors, as well as pay compression ratios in the public and private sectors.

What features of a survey make it eligible for inclusion in the WWBI, and, more broadly, what features are useful for the analytics of public and private labor markets? Since the WWBI focuses on national aggregates, the survey must be representative at the country level (rather than, for example, including just urban areas). Correspondingly, the survey must have taken a sampling approach that attempts to represent each unit of observation across the country equally. Beyond the WWBI, if analysts are only interested in public

sector labor markets in urban areas, the survey should have appropriately sampled within the requisite conurbations of the country.

A second set of requirements relates to the size and composition of surveys included in the WWBI. Specifically, attention is paid to the sample sizes for major categories of respondents. The ability of the WWBI to properly characterize the public and private sector workforces is dependent on the underlying surveys' possessing large enough samples of these two categories of workers that any estimates would approximate the demographics and compensation of the actual labor forces they model. Within the WWBI framework, surveys with fewer than 200 observations for either labor market, or in which either labor market comprises less than 5 percent of all employed individuals within the survey, are excluded from the analysis. More broadly, any survey should be judged on its ability to enable statistically valid inferences on the underlying population.

Third, the survey should have a sufficient sample size for key variables, so as not to be dropped by the WWBI's quality filters. There are three sets of variables we use for the WWBI, presented in table 27.1. If a survey does not include any of the variables shaded in green in table 27.1 or has greater than 40 percent missing/miscoded observations for any of those variables, the survey is disregarded. If a survey is missing any of the variables shaded in blue or has greater than 40 percent missing/miscoded observations for any of those variables, the specific set of variables related to that module is excluded. The gray variables are additional variables that are not universally used in the construction of the WWBI variables, so we do not require them. However, those variables related to sampling are required if sampling weights were used. Finally, the unshaded variables are frequently used to investigate outliers and so are useful to have if available.⁴

The availability of the variables outlined in table 27.1 provides analysts with a basic setup for labor market analysis. Many such analyses look to compare contemporary results over time or across countries. This requires the availability and harmonization of variables across surveys. The WWBI aims to produce statistics that can be compared across time and space and thus faces issues of the classification of employees, the definition of the public sector, and the formulation of wages.

The classification of employed individuals, paid employees, and public paid employees is based on labor and employment status and sector type. Definitions for total, paid, and formal employment are based on the International Labour Organization (ILO) International Classification of Status in Employment (ICSE), making the WWBI and the ILOSTAT databases cross-compatible (fundamental differences in survey coverage, representation, sample size, and timing notwithstanding). According to the ICSE, total employment is defined as

all those of working age who, during a short reference period, were engaged in any activity to produce goods or provide services for pay or profit. They comprise employed persons "at work," i.e., who worked in a job for at least one hour; [and] employed persons "not at work" due to temporary absence from a job, or to working time arrangements (such as shift work, flexitime and compensatory leave for overtime). (ILO 2013, 6)

Paid employment refers to a subsection of total employment and includes only salaried workers, excluding unpaid or own-account (commission-based) employees, employers, and those that are self-employed. Formal employment is a further subset of paid employment and counts those who are employed in formal occupations (possessing a written contract or having access to benefits like health insurance, pensions, or union membership).

A globally harmonized definition of the public sector is hindered due to issues of comparability emerging from the heterogeneous definition of public employees across countries. To avoid this, the WWBI, as a guiding principle, utilizes the more broadly defined *public sector* as opposed to *general government*, as defined by the International Monetary Fund (IMF) *Manual on Government Finance Statistics* (IMF 2014). Specifically, the public sector consists of all institutional units controlled by the central and subnational governments, as well as public corporations that are engaged in market-based activity. Utilizing this broader definition allows for a cleaner comparison across national surveys.

To make wage data as comparable as possible across surveys, the WWBI denotes only the income associated with the occupation used in the analysis (to which the individual dedicated most of their time in the

TABLE 27.1 Variables Required for the Inclusion of a Survey in the WWBI

Metadata variables
Survey ID
Country ID
Year of the survey
Month of the interview
Household ID
Individual ID
Household sampling weight
Strata
Primary sampling units ID
Demographics
Household size
Gender
Age
Urban/rural
Education module application age
Ever attended school
Attending school
Years of education
Level of education (no education, primary, secondary, tertiary)
Labor module
Labor module application age
Labor status <ul style="list-style-type: none"> • Employed • Unemployed • Not in labor force
Employment status <ul style="list-style-type: none"> • Paid employee • Unpaid employee • Employer • Self-employed • Other, workers not classifiable by status
Number of additional jobs
Sector of activity (public vs. private) <ul style="list-style-type: none"> • Public sector, central government, army, NGO, state-owned company • Private
Industry sector classification (minimum one-digit level, but three-digit level is required for occupational decomposition)
Occupational classification (minimum one-digit level, but three-digit level is required for occupational decomposition)
Wage module
Hours of work in the last week
Last wage payment

Source: Original table for this publication, based on World Bank 2021.

Note: The table shows the three sets of variables used. If a survey does not include any of the variables shaded in green or has greater than 40 percent missing/miscoded observations for any of those variables, the survey is disregarded. If a survey is missing any of the variables shaded in blue or has greater than 40 percent missing/miscoded observations for any of those variables, the specific set of variables related to that module is excluded. The gray variables are additional variables that are not universally used in the construction of the WWBI variables, so are not required. ID = identification; NGO = nongovernmental organization; WWBI = Worldwide Bureaucracy Indicators (World Bank).

week preceding the survey) and excludes bonuses, allowances, and other in-cash or in-kind payments from the same job, as well as all additional sources of income (from other jobs) or investments and transfers. Due to the almost complete lack of information on taxes in most household surveys, the wages from the primary job are not net of taxes. For those who are self-employed or own their own businesses, this corresponds to net revenues (net of all costs excluding taxes) or the amount of salary withdrawn from the business.⁵

Wage information in the surveys is reported in each country's local currency units, with a diverse array of periodicity. Great care should be taken to identify the exact frequency of income for each individual within the surveys and convert all wages to a weekly (or another common unit of) wage after accounting for the varying hours worked to ensure credible comparisons across individuals and groups. Additionally, to control for the effect of possibly spurious outliers, the wage variables in the WWBI are winsorized by limiting extreme values in the survey data at the top 0.01 percent level.⁶ More broadly, analysts may want to be cautious with wage information that seems like an outlier from the general distribution of a particular survey.

Overall, to be useful for government analytics, existing household surveys should have sufficient coverage of the population and relevant variables, be of sufficient size, and, where comparisons to international surveys are required, have questions appropriately harmonized with international standards. Fitting these criteria, individual country efforts can always be integrated into existing indicators, such as the WWBI, or compared with relevant surveys in other countries of interest.

Extending Data Collection

What if appropriate household surveys do not exist? Governments, independent organizations, or even individual analysts may be in a position to create and field such surveys. In many instances, project teams from the World Bank have run large, nationally representative household surveys themselves to collect information to aid policy guidance. In India, a private sector organization, the Centre for Monitoring Indian Economy (CMIE), complements the government's labor force survey. By operating the "world's largest household panel survey," with over 2 million individual respondents covering 236,000 households three times a year, the CMIE increases the frequency of up-to-date labor market data for the government and other stakeholders.⁷

Household surveys that are optimized for government analytics could solve the issues with representative sampling identified above by targeting populations of public and private sector workers in a way that ensures an equal probability of inclusion. Such sampling could be done at the subnational level and targeted at those sections of the labor market where the government is particularly prevalent or is aiming to emphasize recruitment. Information not typically collected by household surveys but of substantial interest to those aiming to understand public sector labor markets could be collected, such as information on perceptions of the public recruitment process at different levels of government and how features of the public sector (such as perceived wage and pension benefits) affect respondents' wider labor market choices. Finally, sector variables, such as what specific parts of the government a respondent works in (or its private sector comparator), would allow for analyses that are more precisely targeted at particular job categories.

INSIGHTS EMERGING FROM HOUSEHOLD SURVEYS

Such a systematic utilization of labor force data can allow for the delineation of unique stylized facts on public sector employment and compensation that can provide valuable insights for governments. This section provides illustrative examples of insights into the (relative) nature of government labor markets emerging from household surveys.

The Size of the Public Sector in the Overall Labor Market

A foundation stone of government analytics is the size of the public sector as a share of the national, regional, or local labor market. This topic relates to questions about the appropriate size of government and its impacts on private sector labor markets.

The WWBI reveals that the public sector is a major source of employment in most countries; often, it is the single largest employer. More specifically, the public sector accounts for an average of 16 percent of total employment and over 30 percent and 37 percent of paid and formal employment, respectively. The first metric measures the overall labor market footprint of the public sector, while the latter two are more precise measures of the public sector's relative size within the salaried and formal segments of the labor market.

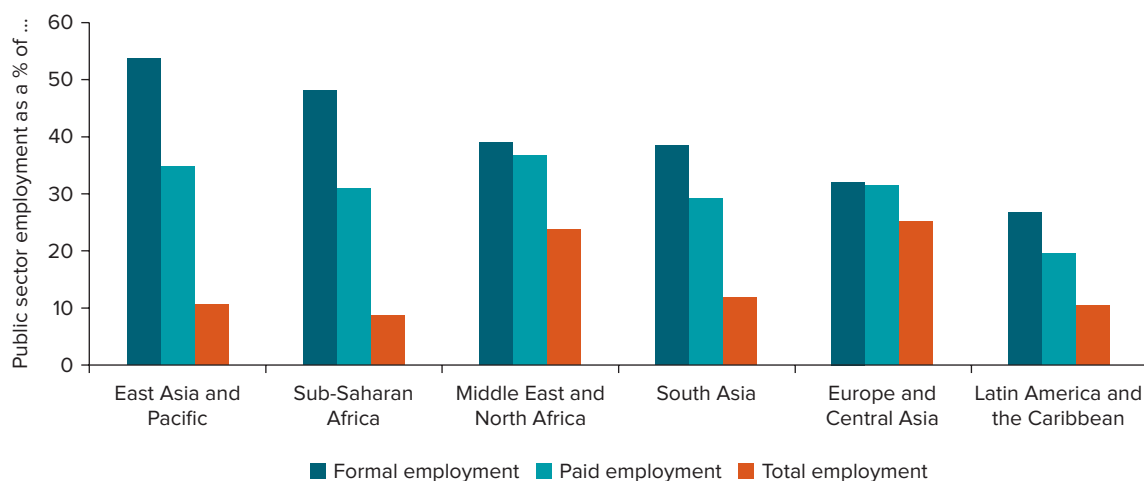
The size and importance of the public sector vary extensively by country income and region (see figure 27.1). While less than 9 percent of the total labor force of the average nation in Sub-Saharan Africa is employed in the public sector, the governments of the Middle East and North Africa employ a quarter of the entire labor force there. This difference is even more stark when looking at formal employment. Such comparisons can be made at the subnational level, allowing the government to develop a sense of how “imposing” its employment is as part of the total stock of formal jobs.

These basic statistics illustrate the wealth of information contained within household surveys that can help governments understand the importance of the public sector, not only as a provider of essential public services but as a key determinant of the health of labor markets, which can help practitioners make better-informed policy decisions.

Further, tracking these indexes over time can help governments understand how the share of the public sector has evolved over time. Figure 27.2 illustrates that, for the world as a whole, a convergence is taking place in terms of the relative size of the public sector. While the public sector's share within total employment has increased, public employment as a share of formal employment has steadily declined over the 18-year period studied.

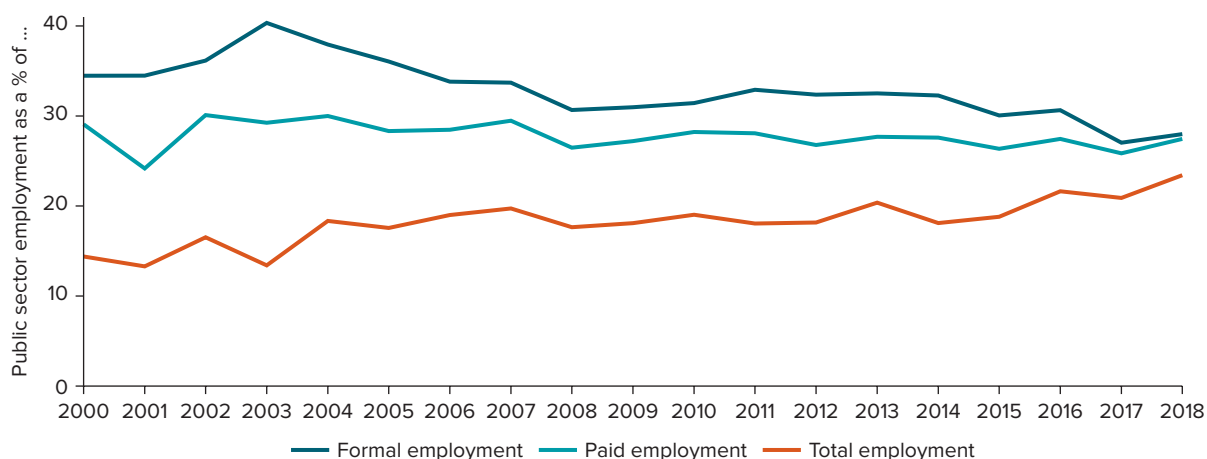
The former is likely due to the fact that as countries develop, their public sectors are called upon to provide more and increasingly complicated services. Conversely, the decline in the public sector's importance within the formal labor force is in part driven by the increased penetration of formal contracting and benefits within the private sector. Regional analysis shows that the relative importance of the public sector within formal employment fell faster and further in middle-income countries than in high- or low-income countries, both of which experienced relatively slower growth rates of labor force productivity and per capita income (Cho et al. 2012).

FIGURE 27.1 Differences in Public Sector Employment, by Region, 2000–18



Source: Worldwide Bureaucracy Indicators, version 2.0, <https://datacatalog.worldbank.org/search/dataset/0038132>.

FIGURE 27.2 Relative Size of the Public Sector Workforce, 2000–18



Source: Worldwide Bureaucracy Indicators, version 2.0, <https://datacatalog.worldbank.org/search/dataset/0038132>.

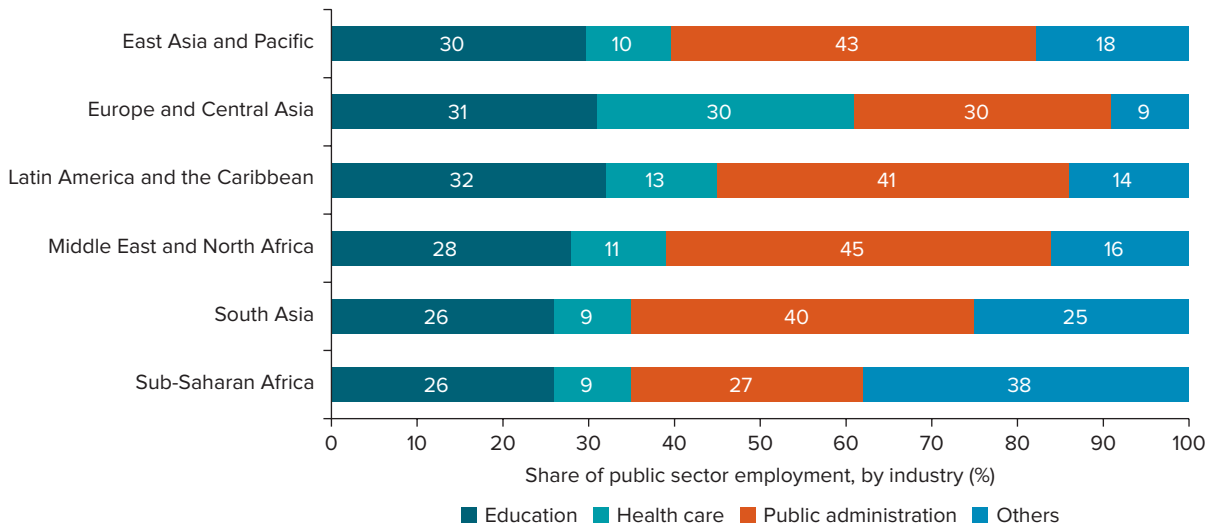
Moreover, household surveys allow policy practitioners to further disaggregate public employment by industry. Countries often have unique legal and occupational classifications for public sector employees, complicating cross-national comparisons. In some countries, all government employees are classified as civil servants, meaning they enjoy distinct legal protections. In others, only management and policy staff are categorized as civil servants, with others, particularly service delivery staff, enjoying fewer privileges and being governed by labor codes similar to private sector employees.

The WWBI reveals that the public administration workforce (which includes individuals responsible for the general administration of the government; the provision of defense, justice, police, and foreign affairs; and the management of compulsory social security) is the single largest segment of the public sector workforce in most countries. On average, 35 percent of the public sector workforce is employed in public administration, followed by the education and health care sectors, which employ, on average, 30 percent and 19 percent of the public sector workforce, respectively. Together, these three industries account for over 80 percent of all public sector employees (figure 27.3). The oversized nature of the health care sector within the Europe and Central Asia region is driven primarily by the extensive public health systems within countries in the European Union.

Additionally, the “other” category in figure 27.3 accounts for public sector employment in all remaining areas of economic activity, ranging from construction and infrastructure to the provision of public utilities, or workers employed within state-owned enterprises other than those involved in public administration, education, or health care provision. Here, countries within the Sub-Saharan Africa region are clear outliers, a phenomenon driven by large public sector penetration in the mining, manufacturing, and services sectors. Given the relatively lower levels of economic development in many countries within the region, this points to the important role that the public sector plays in countries with underdeveloped private sectors. Still, while there may not exist a universal formula for the ideal makeup of the public sector workforce, household surveys can allow a government to benchmark the organization of its public sector workforce across peer countries, or even historically, to track its evolution.

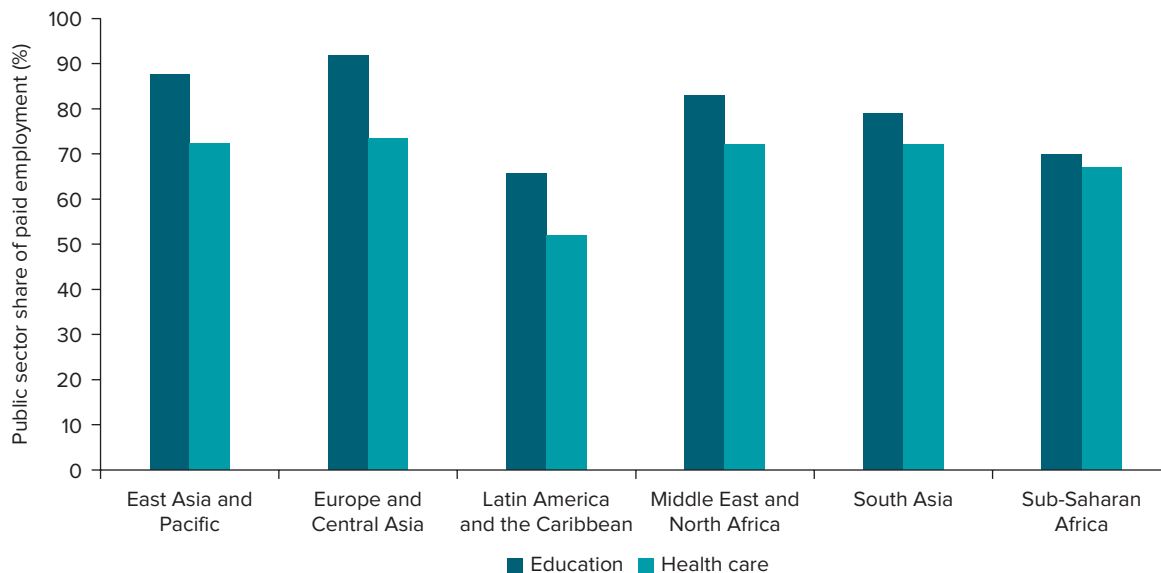
Education and health care workers are essential to a country’s ability to meet its Sustainable Development Goals (SDGs) for the adequacy and universality of health care coverage and education provision. The share of the public sector devoted to the provision of social services differs with country incomes. Looking closely at the education and health care workforce through labor force surveys helps explain the importance of the public sector in the provision of these services. Globally, over three-fourths and two-thirds of the education and health care paid workforce are employed in the public sector, respectively (figure 27.4). This is, in part, driven by the importance that governments across the world place on the provision of education and health care as mandated by the SDGs. It is also partly due to the limited capacity within the private sector to satiate national demand for these services.

FIGURE 27.3 Areas of Economic Activity as a Segment of Public Employment, by Region, 2000–18



Source: Worldwide Bureaucracy Indicators, version 2.0, <https://datacatalog.worldbank.org/search/dataset/0038132>.

FIGURE 27.4 Public Sector Education and Health Care Employment, by Region, 2000–18



Source: Worldwide Bureaucracy Indicators, version 2.0, <https://datacatalog.worldbank.org/search/dataset/0038132>.

Both these segments of the workforce have seen significant attention in the aftermath of the COVID-19 pandemic, as frontline education and health care providers, academics and researchers, epidemiologists, public health experts, and engineers have been an essential bulwark against the public health crisis. Their importance and contribution cannot be overstated. Household surveys can shed light on the role that the public sector education and health care workforce plays within these two sectors. The WWBI finds substantial variation by region (as illustrated in figure 27.4). While over 91 percent of the education workforce and 73 percent of the health care workforce in the Europe and Central Asia region is employed in the public sector, the Latin America and the Caribbean region employs just under 66 percent and 52 percent of these workers, respectively.

The public sector is an important employer for workers with tertiary degrees. Given the particular focus that household surveys place on collecting information systematically on the academic qualifications of workers in the labor force, using globally harmonized measures of educational attainment, these surveys offer a window into the skills makeup of the public and private sector workforces. Looking at data from the WWBI, which tracks the qualifications of workers internationally, we can see that the public sector has a higher proportion of workers with tertiary degrees. Of public sector workers, 47 percent have a tertiary degree, compared to 21 percent in the private sector. (Figure 27.5 provides a dot plot of countries in the WWBI comparing the national shares of tertiary-educated workers in the public and private sectors.) These differences between public and private sector workers have implications for any comparative analysis between the two labor markets, especially public-private wage differentials.

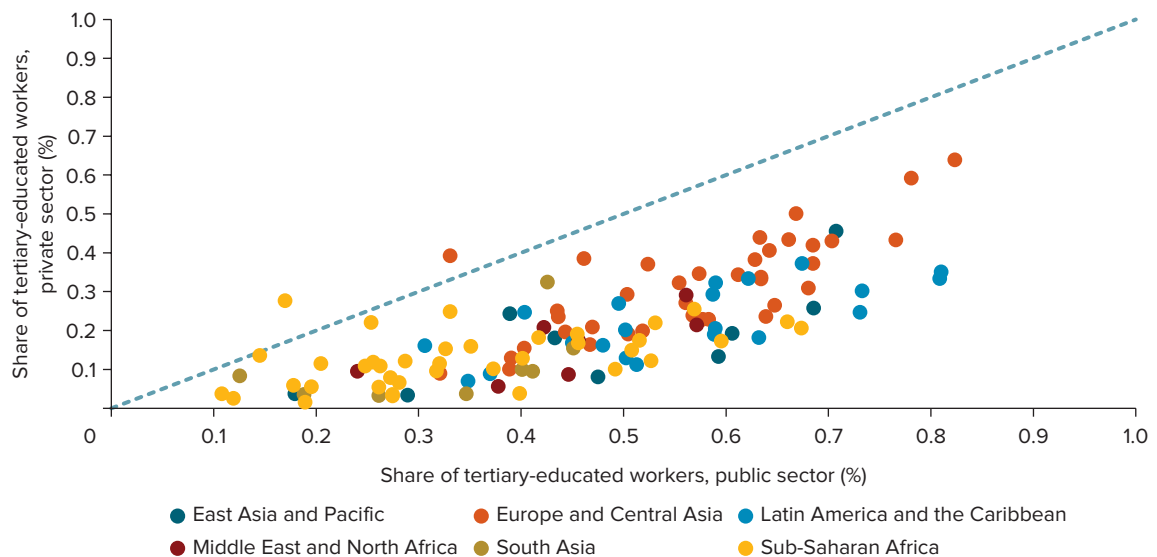
The proportion of public sector workers with tertiary education varies by country income level. In low-income countries, 19 percent of public workers have either no or only primary education, while in high-income countries, this share is negligible. A high proportion of low-skilled workers points to the public sector's serving a social welfare function. A corollary to a high proportion of low-skilled workers is a high proportion of clerical or support jobs. At the other end of the education spectrum, the share of employees with a tertiary degree has increased globally by around 20 percentage points in both sectors over the past decade, but the public sector continues to employ more workers with degrees.

By generating comparative information on the two sectors, possibly over time and across regions and countries, household surveys allow government analysts to understand the broad features of the public sector labor market and the role of the public sector in various national labor markets. A growing body of literature confirms this ability and the importance of the public sector in employing high-skilled workers (Gindling et al. 2020; Grindle and Hilderbrand 1995; Tummers and Knies 2013). Labor force surveys are thus well positioned to enable coherent international comparisons that provide benchmarks to assess a country's current state and dynamics.

Understanding Gender Discrimination

The public sector is an important source of formal employment for women. The public sector's large labor market footprint means that it can be a strategic leader in changing norms and behaviors and promoting

FIGURE 27.5 Tertiary Education among Public and Private Sector Workers, by Region, 2000–18



Source: Worldwide Bureaucracy Indicators, version 2.0, <https://datacatalog.worldbank.org/search/dataset/0038132>.

Note: There is no estimate for the private sector in North America, so the region is not included in the graph.

greater equality in employment in the overall labor market. However, understanding the current state of women’s participation in differential labor markets requires detailed information on the quality of gender representation in the public and private sectors.

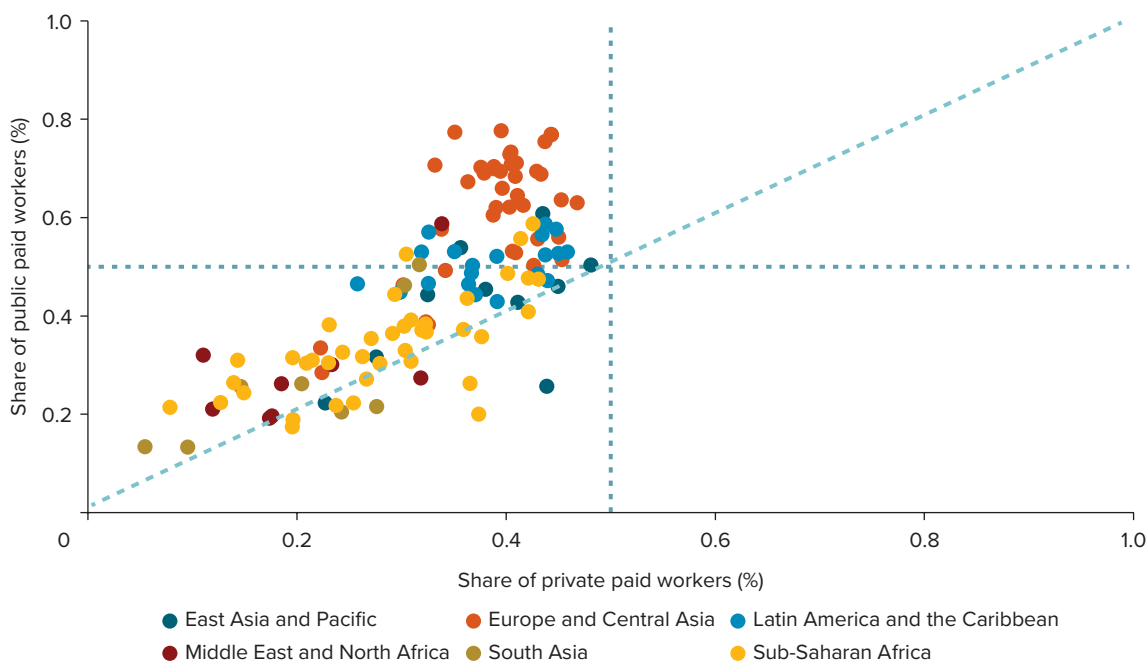
In many developing countries, the public sector, in general, and the education and health care sectors, in particular, have been among the few options for formal employment available to women (Yassin and Langot 2018). Globally, women represent 46 percent of the public sector workforce, compared to 33 percent of the private sector workforce. (Figure 27.6 provides a dot plot of countries in the WWBI comparing the national shares of public and private sector workers who are women.) While men outnumber women in the private sector in all 130 countries for which data are available, women outnumber men in the public sector in 55 countries.

Female representation in the public sector is strongly correlated with country income. A large body of literature finds a U-shaped relationship between female employment in the private sector and economic development (Goldin 1995; Goldin and Polachek 1987).⁸ Labor force surveys included in the WWBI help provide evidence for a positive and significant relationship between female participation in the public workforce and country income. Multiple factors influence female participation rates in the labor force.⁹ A growing body of literature confirms the positive relationship between more-representative bureaucracies (including through female participation) and improved social and economic outcomes across a wide spectrum, including reductions in gender-based violence (Johnston and Houston 2016), improvements in student performance (Zhang 2019), and improvements in public sector productivity (Andrews et al. 2005; Park 2013).

The Appropriateness of Public Sector Wages

Public sector wages are an important determinant of personnel quality and motivation and, therefore, a key determinant of state capacity. However, what is the appropriate level and structure of these wages? Answering this question requires an assessment of who makes up the appropriate comparator group for public sector workers. The first option is to directly compare the wages of public and private sector workers

FIGURE 27.6 Share of Female Workers in the Public versus Private Sector, by Region, 2000–18



Source: Worldwide Bureaucracy Indicators, version 2.0, <https://datacatalog.worldbank.org/search/dataset/0038132>.

Note: The diagonal dashed line represents equity between the two sectors. There is no estimate for the private sector in North America, so the region is not included in the graph.

within a particular country, given that the most likely outside option to employment in the public sector is the corresponding private sector. Estimating public-private wage differentials within a country has been explored in a very large body of academic and policy literature.¹⁰

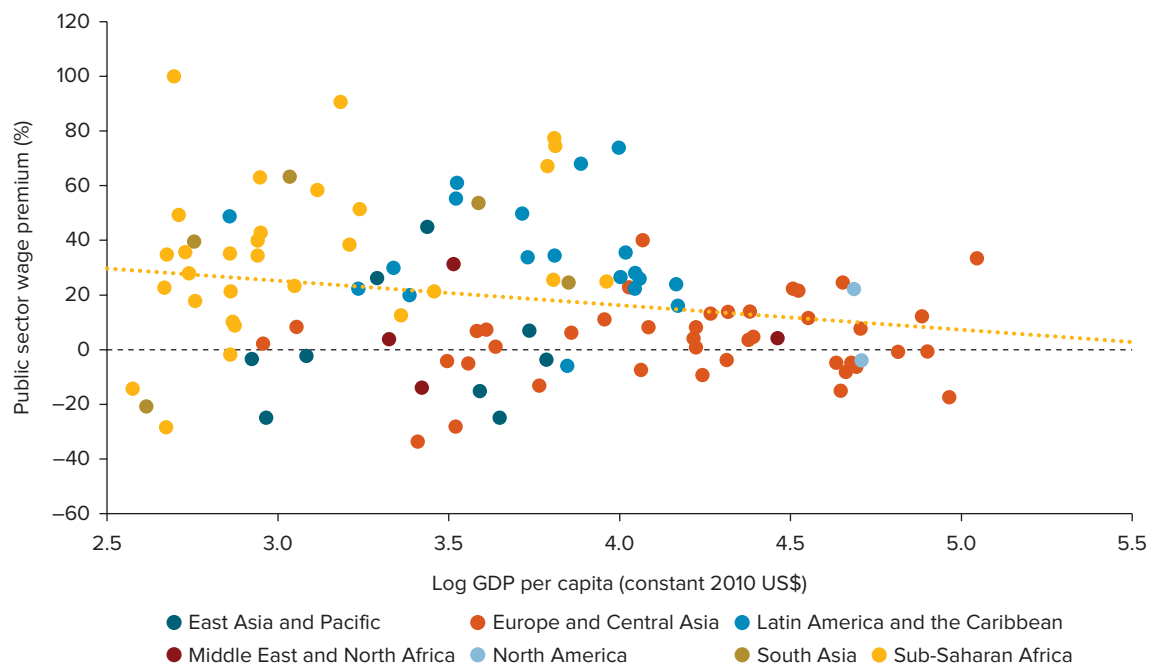
The second approach involves comparing the wages of public sector workers in one country with those of similar workers in other (comparable) countries. Given that these are the closest counterparts to one country's public sector workforce, this is another important method for estimating whether public servants in one country are over- or underpaid. These comparisons are particularly useful in the case of industries or occupations in which workers have transferable skills, such as health care workers who migrate internationally or workers in clerical or managerial positions who rotate within the public sector.

A third option is to compare individuals who perform different tasks or are employed in different occupations within the same country's public sector. This may be useful if public servants are able to move across the service from one organization or region to another.

Household surveys enable each of the above approaches, and such analysis has been undertaken in the WWBI. The data set indicates that public employees in most nations receive a wage premium compared to their counterparts in the private sector. Figure 27.7 shows the premium when the public sector is compared to all private sector salaried employees, irrespective of the type of job and controlling only for worker characteristics (including sex, age, level of education, and location). The figure is ordered by log GDP per capita to provide an indicative sense of premia vary with economic wealth. Public sector workers have approximately 19 percent higher basic wages (excluding allowances and bonus payments) across the 111 countries for which household surveys were sourced, with 80 countries having a positive premium. There is considerable heterogeneity in the size of that premium across countries, varying from a penalty of 33 percent to a premium of 100 percent. The size of the premium is negatively correlated with country income, a finding corroborating academic studies that report higher premiums for developing countries (Finan, Olken, and Pande 2017).

It is important for the government to understand how wage premiums are distributed across worker groups. The public sector wage premium is not uniform and varies by personnel characteristics.

FIGURE 27.7 Public Sector Wage Premium Compared to Country Income, by Region, 2000–18



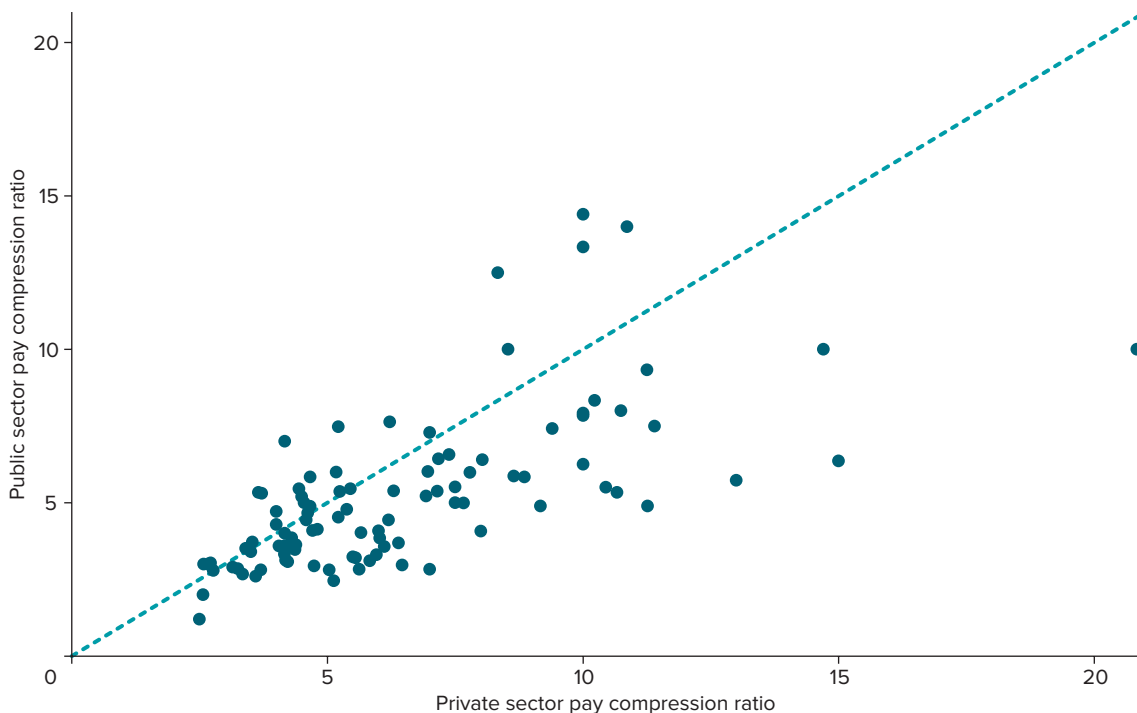
Source: Worldwide Bureaucracy Indicators, version 2.0, <https://datacatalog.worldbank.org/search/dataset/0038132>.

The magnitude of the public sector wage premium depends on an employee’s educational qualifications and is lowest for tertiary-educated officials. The main reason that tertiary-educated individuals earn a low, or no, premium compared to private sector workers is the ability to earn greater wages in the private sector. Similarly, the large wage premium for women in the public sector has greater implications for the large gender pay gaps that exist in the private sector.

While the wage differentials between typical public and private sector workers presented above are worthy of attention from public officials in terms of their impact on the competitiveness of wages in the public sector, the public sector workforce represents a specific subset of the national labor force as employment. Public sector workers are concentrated within a handful of industries (public administration, education, and health care) and certain occupational groups (including managerial, professional, and clerical occupations). Therefore, a second, equally important element of the public sector wage structure for government officials is the difference in wages for workers in different segments of the public sector workforce. Studies have shown that workers compare their wages with their peers in an organization, just as they do with the private sector, and wage differentials that are not perceived as justifiable can be demotivating (Borjas 2002). Additionally, wage equity—whether staff in similar jobs, with similar skills and similar performance, are paid equally—impacts worker motivation and productivity and can be a major driver of the wage bill.

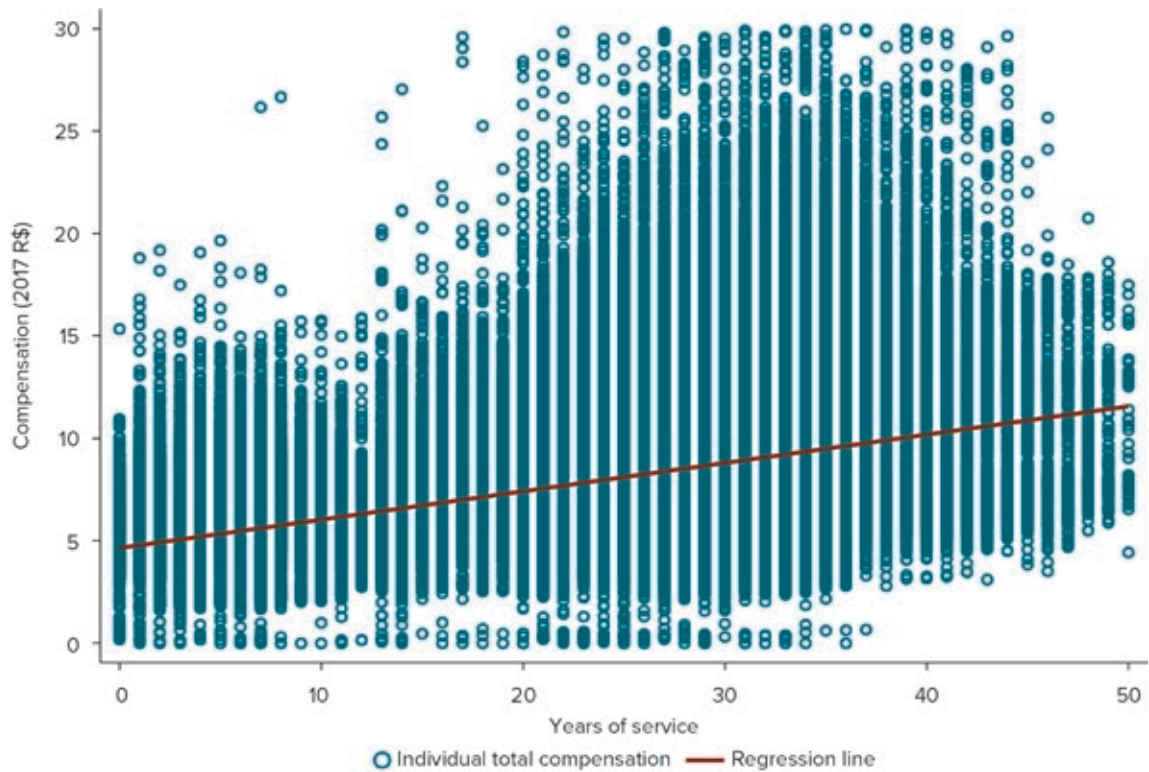
Wage dispersion is generally higher in the private sector than in the public sector. One common metric is the wage compression ratio, which is the ratio of the 90th percentile wage to the 10th percentile wage in the salary distribution. This ratio is lower in the public sector for 70 out of 99 countries for which there are data in the WWBI (figure 27.8). The average wage compression ratio for the public sector across 101 countries is 4.9, compared to 6.3 in the private sector. The lower dispersion in the public sector reveals a trade-off between equity and pay competitiveness at the top of the salary distribution that governments manage. Such information can help public sector managers determine new wage schedules aimed at attracting and maintaining a cadre of high-skilled functionaries in the public sector.

FIGURE 27.8 Public versus Private Sector Pay Compression Ratios, 2000–18



Source: Worldwide Bureaucracy Indicators, version 2.0, <https://datacatalog.worldbank.org/search/dataset/0038132>.
 Note: Each dot represents a country.

FIGURE 27.9 Pay Inequity in the Brazilian Public Sector, 2020



Source: World Bank 2020.

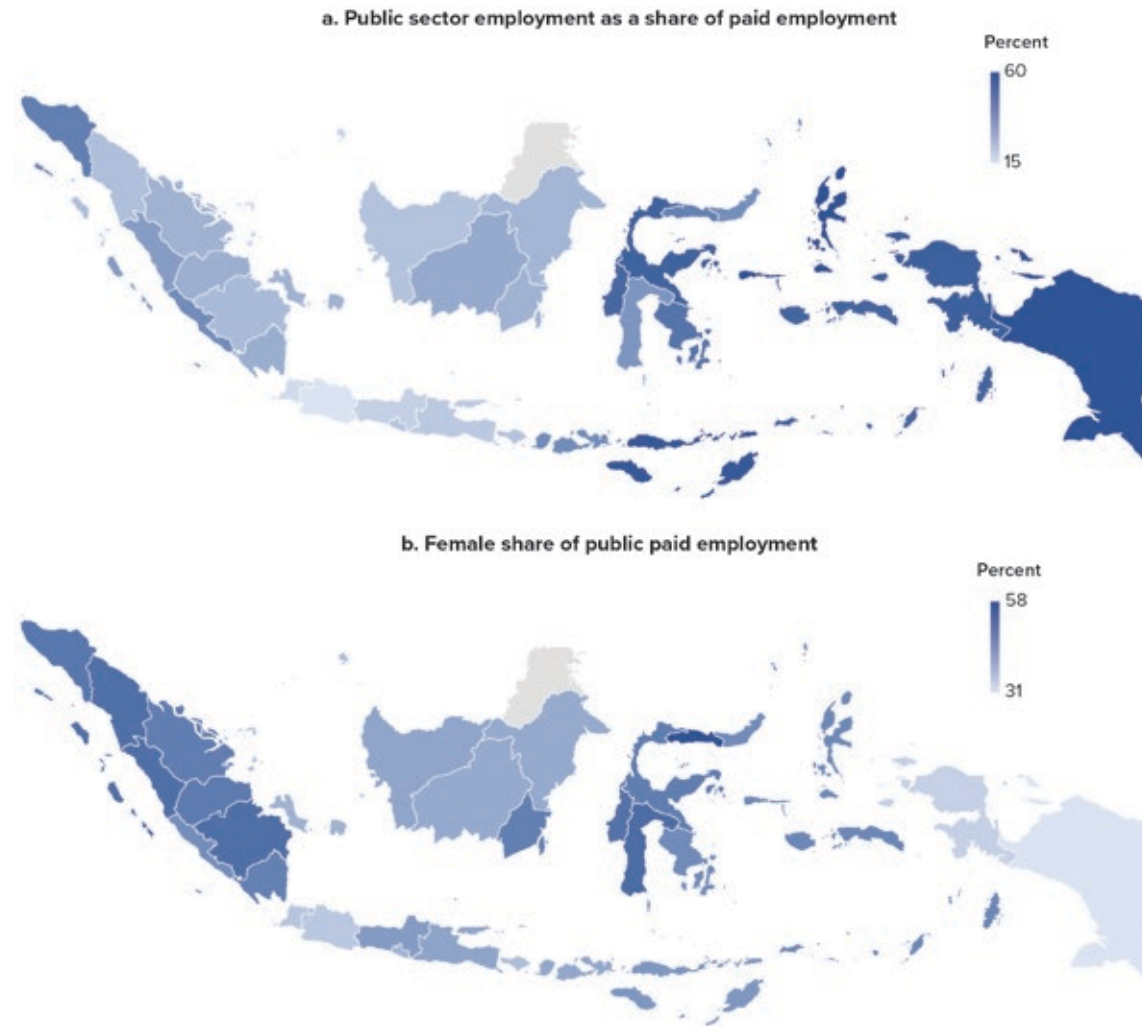
Note: Each dot represents an employee.

Household surveys are also able to provide information on the degree of unexplained variation in wages for individuals employed within similar occupations in the public sector. Figure 27.9 shows that the gross pay received can vary tenfold for workers with similar levels of experience, which is largely a result of non-performance-related payments and not basic pay. While these differences may, in part, be due to personnel demographics (age, gender, or educational qualifications) or the nature of work (job family, industry, or scale), this does point to public sector wages being weakly associated with the experience of workers. Still, these wage differences across employees performing similar tasks and of similar grades but working in different locations or organizations can potentially act as distortions in the workforce.

Understanding Regional Variation

To further illustrate the power of household surveys, if sample sizes are sufficiently large and survey sampling is appropriately stratified, this analytical approach can be replicated at subnational levels. For example, the team has applied the methodology to analyzing public and private sector labor markets at the provincial level in Indonesia. Two variables from this effort are illustrated for Indonesia's provinces in map 27.1. These efforts allow for a closer understanding of regional disparities in the scale, composition, and compensation of the public and private sectors across administrative divisions within countries. In the case of Indonesia, for example, the public sector comprises almost 60 percent of paid employment in the eastern provinces of East Nusa Tenggara, North Maluku, West Papua, and Papua, compared to less than 15 percent in the western provinces (map 27.1, panel a). The female share of public sector employment (which stands at 44 percent at the national level) is mostly concentrated in the eastern and central provinces (map 27.1, panel b). Stylized facts like these can help shed light on many aspects of the nature of public and private labor markets across subnational units within a country.

MAP 27.1 Subnational Patterns in Public Sector Employment, Indonesia, 2018



Source: Original maps for this publication, based on household survey data.

CONCLUSION

We have presented a microdata-based approach for governments to improve their understanding of the public sector workforce and labor markets. Such understanding helps in the development of empirically grounded public service compensation and employment strategies. We have demonstrated how government analysts can use existing household surveys to generate novel insights into government and how these lead to insights that can allow policy makers to make better fiscal choices. Thus, the range of data that should be included for consideration in human resources management information systems, outlined in chapter 9, includes household surveys. Capitalizing on household surveys for government analytics provides a powerful complement to payroll analysis (as described in chapter 10) and broader budget analytics.

These kinds of analytics matter for the effective management of the state, but they also matter for the impact of the public sector on private sector labor markets. Given the size of the public sector, public sector compensation should be designed in cognizance of its influence on the broader labor market. While public sector wage-setting mechanisms do not mechanically respond to market forces, they should be carefully designed to consider the distributional aspects of wages. Policy makers need to ensure that public sector

wages remain competitive enough to attract and retain high-quality public sector workers while not creating disequilibria in private sector labor markets through queuing and crowding effects. Under an optimal compensation policy, public sector wages will be competitive without being distortionary, and there will not be any shortage of skills in either sector.

We have used a series of examples from the World Bank's WWBI to demonstrate how the use of household survey data can help policy makers gain insight into the current and future state of their government's employment and compensation policies. This approach enables researchers, development practitioners, and policy makers to answer some of the most important questions about the appropriate level and distribution of employment in the public sector; the equity, transparency, and market competitiveness of public sector wages; and their impact on fiscal sustainability, the labor market, and service delivery.

NOTES

The approach laid out here leverages the methodological and operational guidelines followed by the Bureaucracy Lab in the construction of the World Bank's Worldwide Bureaucracy Indicators (WWBI), a novel cross-national data set on public and private sector employment and compensation practices. The data set was derived from over 1,000 nationally representative household surveys from 202 countries and territories between 2000 and 2020, providing over 300 granular indicators on the composition, demographics, and compensation of public sector workers. However, this chapter goes beyond that effort to showcase how such an approach can be replicated by researchers, development practitioners, and policy makers to gain a better understanding of the personnel dimensions of state capability, the footprint of the public sector within the overall labor market, and the fiscal implications of the public sector wage bill.

1. There are notable exceptions, such as the Brazilian Ministry of Labor and Employment's *Relação Anual de Informações Sociais* (RAIS) data set, which contains information about employees and businesses for 97 percent of the Brazilian formal market.
2. To counteract these two concerns, governments can limit the presence of these biases in two ways. First, they can ensure that the selection of respondents is based on high-quality census data that guarantee that the sample selected is a good representation of the overall population of the country and, more importantly, is a realistic representation of labor force participants in the public and private sectors. Second, they can look to surveys that include tens of thousands (or an even higher number) of respondents to ensure that any potential weakness in sample selection is alleviated through a large sample size.
3. Further details on the construction of the WWBI are available in a technical note (World Bank 2022). The WWBI data set can be accessed online here: <https://datacatalog.worldbank.org/search/dataset/0038132>. WWBI data are displayed in a dashboard viewable at [https://databank.worldbank.org/source/worldwide-bureaucracy-indicators-\(wwbi\)](https://databank.worldbank.org/source/worldwide-bureaucracy-indicators-(wwbi)). The underlying analytical code has also been made available in the World Bank's repository on GitHub at <https://github.com/worldbank/Worldwide-Bureaucracy-Indicators>.
4. The thresholds used by the WWBI are a product of empirical investigation into the robustness of the indicators to different levels of missingness. More details are provided in the various technical reports accompanying distinct versions of the WWBI (see, for instance, World Bank 2022).
5. Certain surveys do include information on work benefits, such as health insurance and social security, but these are not monetized and cannot be added to wages to provide an estimate of total compensation.
6. *Winsorizing* or *winsorization* is the transformation of statistics by limiting extreme values in the statistical data to reduce the effect of possibly spurious outliers.
7. More information about the CMIE is available on its website, accessible at <https://consumerpyramidsx.cmie.com/>.
8. Female participation is the highest in low-income countries, falling as countries industrialize, and increasing again at high levels of economic development, as the services sector grows.
9. See Jayachandran (2021) for a rich discussion of this literature.
10. See Bales and Rama (2001), Bargain, Etienne, and Melly (2018), Campos et al. (2017), Christophides and Michael (2013), Coppola and Calvo-Gonzalez (2014), Gibson (2009), and Lausev (2014).

REFERENCES

Andrews, Rhys, George A. Boyne, Kenneth J. Meier, Laurence J. O'Toole, and Richard M. Walker. 2005. "Representative Bureaucracy, Organizational Strategy, and Public Service Performance: An Empirical Analysis of English Local Government." *Journal of Public Administration Research and Theory* 15 (4): 489–504.

- Arizti, Pedro, Daniel J. Boyce, Natalia Manuilova, Carlos Sabatino, Roby Senderowitsch, and Erial Vila. 2020. *Building Effective, Accountable, and Inclusive Institutions in Europe and Central Asia: Lessons from the Region*. Washington, DC: World Bank. <http://hdl.handle.net/10986/34031>.
- Baig, Faisal Ali, Xu Han, Zahid Hasnain, and Daniel Rogger. 2021. "Introducing the Worldwide Bureaucracy Indicators: A New Global Dataset on Public Sector Employment and Compensation." *Public Administrative Review* 81 (3): 564–71.
- Bales, Sarah, and Martin Rama. 2001. "Are Public Sector Workers Underpaid? Appropriate Comparators in a Developing Country." Policy Research Working Paper 2747, World Bank, Washington, DC. <http://hdl.handle.net/10986/19338>.
- Bargain, Olivier, Audrey Etienne, and Blaise Melly. 2018. "Public Sector Wage Gaps over the Long-Run: Evidence from Panel Administrative Data." IZA Discussion Paper 11924, Institute of Labor Economics, Bonn, Germany. <https://www.iza.org/publications/dp/11924/public-sector-wage-gaps-over-the-long-run-evidence-from-panel-administrative-data>.
- Behar, Alberto, and Junghwan Mok. 2013. "Does Public-Sector Employment Fully Crowd Out Private-Sector Employment?" IMF Working Paper WP/13/146, International Monetary Fund, Washington, DC.
- Borjas, George J. 2002. "The Wage Structure and the Sorting of Workers into the Public Sector." NBER Working Paper 9313, National Bureau of Economic Research, Cambridge, MA.
- Campos, Maria M., Domenico Depalo, Evangelia Papapetrou, Javier J. Pérez, and Roberto Ramos. 2017. "Understanding the Public Sector Pay Gap." *IZA Journal of Labor Policy* 6 (7). <https://doi.org/10.1186/s40173-017-0086-0>.
- Cho, Yoonyoung, David N. Margolis, David Newhouse, and David A. Robalino. 2012. "Labor Markets in Low and Middle-Income Countries: Trends and Implications for Social Protection and Labor Policies." Social Protection and Labor Discussion Paper 1207, World Bank, Washington, DC.
- Christophides, Louis N., and Maria Michael. 2013. "Exploring the Public-Private Sector Wage Gap in European Countries." *IZA Journal of European Labor Studies* 2 (15). <https://doi.org/10.1186/2193-9012-2-15>.
- Coppola, Andrea, and Oscar Calvo-Gonzalez. 2014. "Higher Wages, Lower Pay: Public vs. Private Sector Compensation in Peru." Policy Research Working Paper 5858, World Bank, Washington, DC. <https://doi.org/10.1596/1813-9450-5858>.
- Finan, F., B. Olken, and R. Pande. 2017. "The Personnel Economics of the Developing State." In *Handbook of Economic Field Experiments*, vol. 2, edited by Abhijit Vinayak Banerjee and Esther Duflo, 467–514. <https://doi.org/10.1016/bs.hefe.2016.08.001>.
- Gibson, John. 2009. "The Public Sector Pay Premium, Compensating Differentials and Unions: Propensity Score Matching Evidence from Australia, Canada, Great Britain and the United States." *Economics Bulletin* 29 (3): 2325–32.
- Gifford, Brian. 2006. "The Camouflaged Safety Net: The U.S. Armed Forces as Welfare State Institution." *Social Politics: International Studies in Gender, State and Society* 13 (3): 372–99.
- Gindling, T. H., Zahid Hasnain, David Newhouse, and Rong Shi. 2020. "Are Public Sector Workers in Developing Countries Overpaid? Evidence from a New Global Dataset." *World Development* 126: 104737. <https://doi.org/10.1016/j.worlddev.2019.104737>.
- Goldin, Claudia. 1995. "The U-Shaped Female Labor Force Function in Economic Development and Economic History." In *Investment in Women's Human Capital*, edited by T. P. Schultz, 61–90. Chicago: University of Chicago Press.
- Goldin, Claudia, and S. Polachek. 1987. "Residual Differences by Sex: Perspectives on the Gender Gap in Earnings." *American Economic Review* 77 (2): 143–51.
- Grindle, Merilee S., and Mary E. Hilderbrand. 1995. "Building Sustainable Capacity in the Public Sector: What Can Be Done?" *Public Administration and Development* 15 (5): 441–63.
- Hasnain, Zahid, Daniel Rogger, John Walker, Kerenssa Mayo Kay, and Rong Shi. 2019. *Innovating Bureaucracy for a More Capable Government*. Washington, DC: World Bank.
- ILO (International Labour Organization). 2013. *Resolution Concerning Statistics of Work, Employment and Labour Underutilization*. Adopted by the 19th International Conference of Labour Statisticians, Geneva, October 2–11, 2013. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_230304.pdf.
- IMF (International Monetary Fund). 2014. *Government Finance Statistics Manual 2014*. Washington, DC: International Monetary Fund. <https://www.imf.org/external/Pubs/FT/GFS/Manual/2014/gfsfinal.pdf>.
- Ingraham, Patricia W., Philip G. Joyce, and Amy Kneeder Donahue. 2003. *Government Performance: Why Management Matters*. Baltimore, MD: Johns Hopkins University Press.
- Jayachandran, Seema. 2021. "Social Norms as a Barrier to Women's Employment in Developing Countries." *IMF Economic Review* 69 (3): 576–95.
- Johnston, Kate, and John Houston. 2016. "Representative Bureaucracy: Does Female Police Leadership Affect Gender-Based Violence Arrests?" *International Review of Administrative Sciences* 84 (1): 3–20.
- Lausev, Jelena. 2014. "What Has 20 Years of Public-Private Pay Gap Literature Told Us? Eastern European Transitioning vs. Developed Economies." *Journal of Economic Surveys* 8 (3): 516–50.
- Moynihán, Donald, and Ivor Beazley. 2016. *Toward Next-Generation Performance Budgeting: Lessons from the Experiences of Seven Reforming Countries*. Washington, DC: World Bank.

- Park, Sanghee. 2013. "Does Gender Matter? The Effect of Gender Representation of Public Bureaucracy on Governmental Performance." *American Review of Public Administration* 43 (2): 221–42.
- Rasul, Imran, and Daniel Rogger. 2018. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *The Economic Journal* 128 (608): 413–46.
- Tummers, Lars G., and Eva Knies. 2013. "Leadership and Meaningful Work in the Public Sector." *Public Administration Review* 73 (6): 859–68.
- World Bank. 2020. "Brazil: Public Sector Wage Bill and Personnel Management." Unpublished manuscript. Washington, DC: World Bank.
- World Bank. 2021. *Worldwide Bureaucracy Indicators Version 2.0: Codebook and Explanatory Note*. Last updated May 26, 2021. Washington, DC: World Bank. <https://datacatalog.worldbank.org/int/search/dataset/0038132/Worldwide%20Bureaucracy%20Indicators?version=2>.
- Yassin, Shaimaa, and François Langot. 2018. "Informality, Public Employment and Employment Protection in Developing Countries." *Journal of Comparative Economics* 46 (1): 326–48.
- Zhang, Youlang. 2019. "Representative Bureaucracy, Gender Congruence, and Student Performance in China." *International Public Management Journal* 22 (2): 321–42.

CHAPTER 28

Government Analytics Using Citizen Surveys

Lessons from the OECD Trust Survey

Monica Brezzi and Santiago González

SUMMARY

This chapter takes stock of the work of the Organisation for Economic Co-operation and Development (OECD) on defining and measuring what drives people's trust in public institutions. It presents an updated framework on the public governance determinants of trust, demonstrating that competence (for example, responsiveness and reliability) and values (for example, openness, integrity, and fairness) are fundamental public governance levers to improve or harm levels of trust in different institutions, while also recognizing the importance of people's perception of how intergenerational and global challenges are handled by institutions, as well as cultural, economic, and political drivers. In addition, it shows that it is possible, on the basis of this framework, to advance in gathering citizens' evaluations of the functioning of governments and public governance by designing nationally representative population surveys with high standards of statistical quality. The OECD Survey on Drivers of Trust in Public Institutions (OECD Trust Survey) has been specifically designed and tested to capture people's expectations of and experiences with public institutions around the five drivers of trust while controlling for socioeconomic, political, and institutional characteristics. In a handful of countries that, in addition, have conducted in-depth case studies, the evidence resulting from the survey has proven to be a key input for improving policy making, leading to concrete actions for building or strengthening institutional trust. The inaugural cross-country survey was applied to 22 OECD countries at the end of 2021, and the results were made available in June 2022. Consolidating this evidence will be essential to enhancing benchmarking and monitoring of the evolution of policies over time.

Monica Brezzi is head of the Territorial Analysis and Statistics Unit and Santiago González is an economist/policy analyst at the Organisation for Economic Co-operation and Development (OECD) Directorate for Public Governance.

ANALYTICS IN PRACTICE

- It is possible to measure trust and its drivers through household surveys and to produce actionable evidence. Trust indicators have been widely collected by unofficial data producers and are commonly picked up by the media as metrics of government performance. However, indicators on the public governance drivers of trust have been scarcely and inconsistently produced and are not commonly found in population surveys. Over the past years, a body of theoretical and empirical evidence has consolidated, leading to the development of the Organisation for Economic Co-operation and Development Survey on Drivers of Trust in Public Institutions (OECD Trust Survey) as an effort to fill this gap and develop comparative statistics on what drives trust in public institutions. The survey's unique focus on the drivers of trust aims to provide countries with actionable evidence on the strengths and weaknesses of governments and public institutions.
- Bringing together the statistical and policy communities is a key step for ensuring the production of statistics with high standards of quality. Other than their potential sensitivity, there are no good reasons why statistics on trust and its drivers cannot be subject to the same quality standards and requirements that apply to other social, economic, and environmental statistics. The development of the OECD Trust Survey has been led by the OECD Public Governance Committee and has brought together policy makers at the central level of government and representatives from national statistical offices (NSOs), who have reviewed the process for developing and implementing the survey. Bringing together both communities has the advantage of reconciling the development of meaningful statistics with adherence to high requirements and standards of quality. While in most countries, the OECD Trust Survey has been implemented by the OECD via a survey provider, in Finland, Ireland, Mexico, and the United Kingdom, it has been implemented by NSOs.
- The OECD competence-values framework has proven to be a powerful analytical tool to understand the public governance drivers of trust in institutions and provide evidence to public administrations on how to increase their effectiveness to enhance trust. Following the COVID-19 pandemic and the resulting social and economic crisis, a consultative process to revisit the framework on the drivers of trust in public institutions was launched. The process resulted in the enlargement of the framework by adding a section on the perception of public institutions' action on intergenerational and global challenges and further recognition of individual cultural, economic, and political drivers. However, the building blocks of the framework—competence and values—have proven robust, offering a compelling and encompassing framework for understanding and measuring public administrations' work and how they could advance in building trust.
- The drivers of trust vary by institutional actor and level, calling for different types of actions to influence them. The debate on public trust has been dominated by measures of *trust in government* that, to a large extent, are proxies of political trust or trust in the current government. Yet distinctions by the level of government (for example, local government) or between “politicians” and “bureaucrats” have not been rigorously considered. The OECD Trust Survey considers differences across institutions and levels of government and allows for a differentiated analysis of what drives each of them.
- Institutional trust remains a multidimensional construct: drivers and actions to preserve it or restore it are also influenced by context. In addition to the competence and values of public institutions, trust is also influenced by cultural and socioeconomic determinants. Accordingly, baseline levels of trust in a given country are influenced by its history and the current moment. Because cross-country comparisons of trust levels are informative and appealing, contextual elements should be considered when interpreting data and carrying out comparisons. In-depth analysis combining quantitative and qualitative methods at the country level is an important complement to advancing actions for restoring trust.

- Evidence-based regional dialogue and experience-sharing could be powerful tools for strengthening institutional trust. The potential effect of cultural context can be minimized by comparing countries with similar histories, institutional settings, and contexts. Accordingly, evidence generated by the OECD Trust Survey could be a powerful driver of regional dialogue and experience-sharing about actions that could influence public trust in similar contexts.

INTRODUCTION

The availability of internationally comparable data on how public administrations function and perform has dramatically improved over the past 10 years (OECD 2021b). However, measures of governance outcomes, which can help benchmark the effectiveness of public governance, are still scarce. The reasons behind this are conceptual and methodological in nature. On the conceptual side, measuring governance outcomes implies consideration of democratic standards that go beyond a performance-oriented notion of public management. Methodologically, public surveys of an international scope raise questions regarding the validity and robustness of these measures.

Measurement of governance outcomes includes, among other things, assessing levels of satisfaction with public services or *trust in government*. The inclusion in the United Nations' 2030 Agenda for Sustainable Development of a *governance* Sustainable Development Goal (SDG16) has given new impetus to the development of metrics for the outcomes or results of political institutions and processes and to the expansion of household surveys to measure governance results (United Nations Praia Group 2020). In particular, the United Nations Praia Group on Governance Statistics' *Handbook on Governance Statistics* takes stock of existing practices in governance data collection along eight dimensions of public governance, discusses the steps required to achieve international statistical standards, and, in some cases, proposes common questionnaires to be included in household surveys.

This chapter discusses how population surveys can be used to measure governance outcomes and provide guidance to public administrations on actions to increase their effectiveness. After some examples of existing national citizen surveys, the chapter presents recent developments in measuring people's trust in government and its main determinants. Measures of people's trust in government are commonly used indicators of public administration performance (or results) and are included as one of the eight dimensions of the *Handbook on Governance Statistics*.

This chapter presents results from the application of the Organisation for Economic Co-operation and Development Survey on Drivers of Trust in Public Institutions (OECD Trust Survey) as a key example to better understand opportunities and challenges in using citizen surveys to develop evidence on governance outcomes at a national and international scale. The OECD Trust Survey builds on an analytical framework structured along five key dimensions: responsiveness, reliability, openness, integrity, and fairness. Preliminary results suggest important insights from the survey. We find that different public institutions, from national to local governments, map onto different drivers of public institutional trust. Survey findings can guide governments in prioritizing specific areas to strengthen institutional trust.¹

USING CITIZEN SURVEYS TO MEASURE GOVERNANCE OUTCOMES

Governance statistics are fundamental to ensuring that the relationship between the state and its people is inclusive, transparent, and accountable. However, while public governance issues figure prominently in national and international policy discussions, the state of available statistics (in terms of international standards, measurement guidelines, and the availability of comparable information from official sources)

lags well behind the quality of information available in other fields (for example, economic, social, and environmental statistics). As detailed in chapter 27 of *The Government Analytics Handbook*, the World Bank's Worldwide Bureaucracy Indicators are a major exception, aggregating household survey data to measure key public administration productivity inputs (such as public sector wage premiums).

Traditionally, there are two main limitations of public governance statistics. First, most measures refer to the input and processes of governments, while more limited evidence is available on the outputs of governments and the outcomes of public governance, such as the assessment of and satisfaction with public services or trust in government. Second, most governance measures are generated through expert assessments aimed at capturing how governments work. These assessments

typically refer to professionals such as academics, lawyers or civil servants who answer a questionnaire in the area of interest. The resort to experts is thought to be advantageous because they can assess complex topics related to governance through an informed judgment. Expert assessments are frequently combined to create composite governance indexes that summarize multi-concept and complex phenomena into rankings and reference data. There are, however, reasonable concerns on how representative a sample of experts is of the universe of people with knowledge in the matter of interest, as well as about the degree of validity and reliability of data obtained through expert assessments. (United Nations Praia Group 2020, 15; see also González, Fleischer, and Mira d'Ercole 2017; Kaufmann and Kraay 2007)

Broadening the measurement approach to also include citizen surveys can help address both limitations, by shifting the focus to the outcomes of public governance as perceived and experienced by people and by including people's views, through nationally representative population samples, in addition to expert assessments (Fukuyama and Recanatini 2018).

Citizen surveys can also strengthen government accountability. When participating in regular population surveys, citizens are invited to provide feedback on different aspects of public governance, which allows governments to gather input and hear the people's voice beyond electoral processes. Governments and policy makers can also use survey results to better inform policies, identify citizens' priorities and concerns, and assess the support for or impact of different initiatives.

Over the past few years, some developments at international levels, which build on national experiences, have helped to increase the availability of governance statistics. The inclusion of a *governance* goal in the United Nations' 2030 Agenda for Sustainable Development, as mentioned above, and the agreement of the Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs) on several indicators for global reporting in this field (which creates reporting obligations for national statistical offices, or NSOs) are expected to increase the demand for comparable, high-quality evidence on public governance and to broaden measurement approaches. For over a decade, in its flagship report *Government at a Glance*, the OECD has published internationally comparable data on governments' inputs, activities, outputs, and, to the extent possible, outcomes in OECD countries.

The indicators feeding these publications focus on how governments perform from an international perspective, allowing countries to benchmark their performance relative to other countries and over time and providing evidence to policy makers about areas where further progress is needed. Furthermore, initiatives such as the OECD Guidelines on Measuring Trust (OECD 2017a, 2017b) and the Praia Group's *Handbook on Governance Statistics* (United Nations Praia Group 2020) have provided methodological guidance to NSOs on how to develop comparable measures for several dimensions of public governance.

Specifically, the Praia Group, set up in 2014, has mapped in its *Handbook on Governance Statistics* existing measurement initiatives on eight dimensions of public governance: nondiscrimination and equality, participation, openness, access to and quality of justice, responsiveness (satisfaction with services and political efficacy), absence of corruption, trust, and safety and security. The *Handbook on Governance Statistics* highlights the paucity of statistical standards, technical guidelines, and methodological tools that currently exist in many dimensions of governance statistics (for example, on

discrimination, participation in political and public affairs, openness, access to civil justice, satisfaction with services, and forms of corruption other than petty bribery); at the same time, it sheds light on the feasibility of generating comparative evidence on some dimensions of public governance through population surveys, notably on crime victimization, access to criminal justice, political efficacy, and trust in public institutions (United Nations Praia Group 2020).

The *Handbook on Governance Statistics* also highlights some methodological challenges of household surveys—in particular, respondent burden and accessibility standards for specific populations (for example, people with disabilities)—that should be addressed in questionnaire design as well as survey method selection. The authors note that

with a sufficient sample size, survey results may be representative of the population of a country, of a specific area (such as a province or a small locality) or a specific group (e.g., urban/rural populations). However, such disaggregations for governance statistics are not commonly found in national statistical systems and will require special attention from the statistical community (United Nations Praia Group 2020, 14).

Nevertheless, they acknowledge that “household surveys are particularly useful to give voice to marginalized populations in contexts in which the mechanisms to respond to citizens’ demands are not yet consolidated” (United Nations Praia Group 2020, 14; see also AU and UNDP 2017).

One public governance area in which population surveys would provide relevant information to governments is experience and satisfaction with public services. More countries are putting in place regular surveys on public services, although with several challenges. For example, since 2010, the Agency for Public Management and eGovernment in Norway has carried out the Norwegian Citizen Survey. The survey provides a substantial knowledge base for assessing the performance of public services across different sectors and levels of government. The survey is understood as an additional way to engage citizens apart from direct mechanisms, and it addresses perceptions, expectations, and areas of improvement, aiming to develop public outputs and services in a more user-friendly manner, based on citizens’ needs and expectations.

However, the evolution of the survey over time has shown that respondents may find it difficult to answer “customer satisfaction” questions on a large number of services. Since 2018, the survey has included only the largest state-provided services (tax administration, hospitals, police, etc.) and services at the municipal level to which people have regular access (for example, schools). The survey also asks about recent experiences with these services as one criterion that could shape satisfaction levels. The questions address the quality and accessibility of services; satisfaction with information, communication, and consumer orientation; perceived competence, capacity, and trust in the public authority; and general satisfaction (OECD 2022b).

Similarly, the Citizen Experience Survey, conducted since 2019 by the Department of the Prime Minister and Cabinet in Australia, measures public satisfaction, trust, and experiences with Australian public services, with the aim of improving public services and making them more citizen-centered. The survey aims to provide the right evidence for governmental agencies to act on: when data flag “hot spots” in the system that warrant extra attention, senior policy makers and other stakeholders engage in dialogue to identify actions that can help improve services.²

Surveys of the general population can also provide useful data about citizens’ perceptions of the quality of governance in the country where they live, an outcome that is usually measured through expert assessments. The General Population Poll (GPP) by the World Justice Project (WJP) asks citizens about their perceptions of and experience with public institutions.³ The survey inquires about a wide range of topics, from access to public services (including access to information and justice), respect for the law by private and public actors (for example, abuses of power), and civic participation. Data coming from the GPP are included in the WJP Rule of Law Index, which ranks countries according to the quality of their governance, alongside the results of an expert assessment.

WHY MEASURE PUBLIC TRUST?

People's trust in public institutions has long been considered a key outcome of good governance (OECD 2021b). Accordingly, institutional trust happens when citizens consider the government and its institutions in general, as well as individual political leaders, to be promise-keeping, efficient, fair, and honest (Blind 2007). In this chapter, *trust* is defined as a person's belief that another person or institution will act consistently with expectations of positive behavior (OECD 2017a, 2017b). While there are several trust relationships, this chapter focuses on the interaction between governments and citizens, or *institutional trust*, which refers to people's appraisal of public institutions.

Institutional trust is recognized as an important foundation upon which the legitimacy and sustainability of political systems are built. Trust is the basis of the social contract that allows for the delegation of power and sets the basis for democracies to work. In turn, trust is essential for social cohesion and well-being because it affects a government's ability to govern and enables it to act without having to resort to coercion. The COVID-19 pandemic has underlined the importance of public trust for enhancing and accepting behavioral change in order to achieve a collective objective, as well as its importance for achieving compliance (Bargain and Aminjonov 2020). Recent research has shown that support for future-oriented policies on global challenges, such as climate change, is mediated by people's institutional trust (Fairbrother et al. 2021). In a low-trust context, citizens will prioritize immediate, appropriable, and partial benefits and will induce politicians to seek short-term and opportunistic gains through free-riding and populist attitudes (Gyórfy 2013).

Still, institutional trust remains a complex, multidimensional concept influenced by a wide array of facts, circumstances, experiences, and perceptions (OECD 2013). Trust metrics are often quoted by the press and have the capacity to raise awareness and trigger institutional reactions; however, they remain poorly understood. Furthermore, it is not always clear what lies behind these signals and to what extent it is possible to influence them. The OECD's work has focused on conceptualizing the main determinants of institutional trust and developing comparative evidence on them based on citizen surveys.

The remainder of this chapter explores existing metrics of institutional trust, presents the OECD policy framework and its accompanying measurement strategy for the public governance determinants of institutional trust, and concludes with lessons that could inform a measurement agenda moving forward.

EXISTING METRICS FOR INSTITUTIONAL TRUST

In some contexts, there is a long tradition of collecting metrics of trust in government. For instance, in the United States, the Pew Research Center has measured "government confidence" since at least 1958. Similarly, the American National Election Study (ANES), a project of the Center for Political Studies at the University of Michigan, has collected survey-based measures of trust, associated with electoral cycles, since at least 1952. However, it is only since the beginning of the 21st century that cross-country comparative statistics of institutional trust have become widely and regularly available. González and Smith (2017) review these metrics and find seven cross-country comparative surveys (commercial and noncommercial) that have regularly collected trust data since 2002. These surveys have different coverage periodicity and work under different criteria of statistical quality. The most widely used of these surveys is the Gallup World Poll because of its extensive country coverage, its time extension, and the annual frequency of its data.

However, existing measures of institutional trust have shortcomings. Some are technical (such as sampling, scale, and level of representativeness for some population groups); others are conceptual (such as the meaning of *government*). In this chapter, we use *institutional trust* to indicate trust in different types of institutions, such as trust in political institutions (for example, the parliament), trust in administrative

institutions (for example, the civil service or public administration) and nongovernmental institutions (for example, the media), and trust in justice and law administration (for example, the police). Accordingly, institutional trust is measured using a general formulation: “Do you have confidence in ...?” or “How much do you trust ...?” followed by a detailed list of institutions. Empirical analysis suggests that, given the wide range of institutions, people’s responses can be grouped into three categories: political and administrative institutions, law-and-order institutions, and nongovernmental institutions (González and Smith 2017).

Based on existing evidence, the OECD has developed the Guidelines on Measuring Trust (OECD 2017a, 2017b). The guidelines mark an advance in providing an analysis of the accuracy of trust measures. Accuracy has two dimensions: *reliability* and *validity*. The *reliability* of a metric is the degree to which repeated measurements of the same thing produce the same results. In this sense, a reliable measure involves minimal *noise*, or random errors in the measurement process. *Validity* is usually analyzed in terms of *face validity* (whether the measure makes sense intuitively), *convergent validity* (whether the measure correlates well with other proxy measures of the same concept), and *construct validity* (whether the measure behaves as theory and common sense dictate).⁴ The guidelines have found good evidence on the reliability and construct validity of existing trust metrics, but evidence on their face and convergent validity is scarce and inconclusive, calling for further research in the area. The guidelines also propose modules for measuring institutional trust, including an experimental module on the determinants of public trust that will be discussed later in this chapter.

MEASURING WHAT DRIVES PUBLIC TRUST IN INSTITUTIONS

The OECD Framework on Drivers of Trust in Public Institutions

At least three trends emerge from the academic literature for understanding what drives levels of trust in institutions. One theory emphasizes the role of culture and argues that individuals learn to trust or distrust based on early socialization and interpersonal networks, which, in turn, influence their trust in institutions (Tabellini 2008). A second stream of work recognizes the importance of the economic cycle, as well as economic and personal characteristics and preferences (Algan et al. 2018, 2019). Finally, institutional theories focus on the performance and reputation of institutions, both in terms of processes and outcomes, as the key determinants explaining levels of institutional trust (Bouckaert 2012; Rothstein 2013; Van de Walle and Migchelbrink 2020). While institutional trust is probably influenced by a combination of elements driven by culture, economic conditions, and institutions, the OECD’s work on understanding drivers of trust in public institutions has, since 2013, emphasized the importance of high-performing institutions for building public trust.

Understanding the effects of institutions on trust depends on the congruence of people’s preferences (their interpretations of what is right and fair and what is unfair) and their perceptions of the actual functioning of government (Van de Walle and Bouckaert 2003). Other authors have distinguished between *trust in competence*, the ability to deliver on expectations, and *trust in intentions*, performing in good faith according to one’s competence (Nooteboom 2006). These distinctions are extended by Choi and Kim (2012) and Bouckaert (2012), who distinguish between the *logic of consequences*, where trust is derived causally from outcomes, and the *logic of appropriateness*, where trust is based on values, such as integrity and transparency.

Despite the complexity of the subject and the variety of approaches, there is consistency across the literature on institutional trust in at least two key aspects. First, the literature highlights two different but complementary components that matter in understanding and analyzing trust: *competence*—operational efficiency, or the ability, capacity, and good judgment to actually deliver on a given mandate—and *values*—the underlying intentions and principles that guide actions and behaviors. Second, there is consistency in the literature regarding specific attributes that matter for trust, in relation to both competence and values (see table 28.1).

TABLE 28.1 Deconstructing Citizens’ Trust in Public Institutions

Trust component	Government mandate	Concerns affecting trust	Policy dimension
Competence: <i>The ability of governments to deliver to citizens the services they need, at the quality level they expect</i>	Provide public services	<ul style="list-style-type: none"> • Access to public services, regardless of socioeconomic condition • Quality and timeliness of public services • Respect in public service provision, including response to citizens’ feedback 	Responsiveness
	Anticipate change and protect citizens	<ul style="list-style-type: none"> • Anticipation and adequate assessment of citizens’ evolving needs and challenges • Consistent and predictable behavior • Effective management of social, economic, and political uncertainty 	Reliability
Values: <i>The drivers and principles that inform and guide government action</i>	Use power and public resources ethically	<ul style="list-style-type: none"> • High standards of behavior • Commitment against corruption • Accountability 	Integrity
	Inform, consult, and listen to citizens	<ul style="list-style-type: none"> • Ability to know and understand what government is up to • Engagement opportunities that lead to tangible results 	Openness
	Improve socioeconomic conditions for all	<ul style="list-style-type: none"> • Pursuit of socioeconomic progress for society at large • Consistent treatment of citizens and businesses (vs. fear of capture) 	Fairness

Source: Original table for this publication.

Building on the above, the OECD has put forward an analytical framework that offers an instrumental approach to building citizens’ trust in public institutions, facilitating measurement efforts (both based on experience and expectations) and policy attempts to influence trust. The OECD Framework on Drivers of Trust in Public Institutions, developed in 2017 and reviewed in 2021 through broad consultation with academics, policy makers, and civil society, includes four components (Brezzi et al. 2021). The updated framework is presented in table 28.2.

First, the framework places a greater emphasis on capturing trust levels in a larger set of institutions—for example, political parties or intergovernmental organizations—to further recognize the variety of institutions that influence policy making and that can shape people’s assessment of public affairs as well as leaders’ behavior. In addition, it recognizes the importance of improving the representation of diverse population groups that may be systematically excluded from voicing their views in traditional democratic processes, either due to personal characteristics (for example, geography or socioeconomic background) or because they persistently distrust “the system” and opt out of opportunities to express their voice.⁵ The two broad dimensions of public sector competence and values—disentangled in responsiveness, reliability, integrity, openness, and fairness—remain core to the framework, as tested through country studies in the Republic of Korea, Finland, and Norway (OECD 2021a, 2022b; OECD and KDI 2018).

Third, the revised framework presents an “overlay” of the cultural, political, and economic factors that, at both an individual and group level, strongly influence levels of trust in government. Institutional competence and values are, in fact, mediated by individual and group identities, traits, and preferences—including political attitudes. These revisions attempt to emphasize more strongly the role played by political attitudes, including disengagement with the system, in explaining institutional trust.

Finally, the revised framework underlines the role people’s confidence plays in the sustainability and effectiveness of policy action to address long-term and global challenges (for example, climate change, fiscal sustainability, digitalization, and inequality) (Brezzi et al. 2021). As the issues tackled by public institutions become increasingly complex, with long-term consequences involving a larger set of governmental and non-governmental actors, greater coordination and the ability of institutions to manage uncertainty and address trade-offs (for example, generational and economic trade-offs) will be key to preserving social cohesion and maintaining institutional trust.

TABLE 28.2 OECD Framework on Drivers of Trust in Public Institutions

Levels of trust in different public institutions		
Trust in national government, local government, civil service, parliament, police, political parties, courts, legal systems, and intergovernmental organizations		
Public governance drivers of trust in public institutions		
Competence	Responsiveness	<ul style="list-style-type: none"> • Provide efficient, quality, affordable, timely, and citizen-centered public services that are coordinated across levels of government and satisfy users • Develop an innovative and efficient civil service that responds to user needs
	Reliability	<ul style="list-style-type: none"> • Anticipate needs and assess evolving challenges • Minimize uncertainty in the economic, social, and political environment • Effectively commit to future-oriented policies and cooperate with stakeholders on global challenges
Values	Openness	<ul style="list-style-type: none"> • Provide open and accessible information so the public better understands what the government is doing • Consult, listen, and respond to stakeholders, including through citizen participation and engagement opportunities that lead to tangible results • Ensure equal opportunities to participate in the institutions of representative democracy
	Integrity	<ul style="list-style-type: none"> • Align public institutions with ethical values, principles, and norms to safeguard the public interest. • Make decisions and use public resources ethically, promoting the public interest over private interests, while combatting corruption • Ensure accountability mechanisms between public institutions at all levels of governance • Promote a neutral civil service, whose values and standards of conduct prioritize the public interest
	Fairness	<ul style="list-style-type: none"> • Improve living conditions for all • Provide consistent treatment of businesses and people regardless of their background and identify (for example, gender, socioeconomic status, racial/ethnic origin)
Cultural, economic, and political drivers of trust in public institutions		
<ul style="list-style-type: none"> • Individual and group identities, traits, and preferences, including socioeconomic status and interpersonal socialization and networks • Distrust toward and disengagement with the system 		
Perception of government action on intergenerational and global challenges		
<ul style="list-style-type: none"> • Perceptions of government commitment and effectiveness to address long-term challenges 		

Source: Brezzi et al. 2021.

Note: OECD = Organisation for Economic Co-operation and Development.

The Measurement Strategy

The OECD Framework on Drivers of Trust in Public Institutions is operationalized through a nationally representative population survey: the OECD Trust Survey. An experimental module of questions to measure the five public governance drivers of trust was included in the OECD Guidelines on Measuring Trust (OECD 2017a, 2017b). The statistical feasibility and empirical relevance of the population survey were tested in six countries (France, Germany, Italy, Slovenia, the United Kingdom, and the United States) through the OECD Trustlab in 2018 (Murtin et al. 2018) and in Korea, Finland, and Norway through in-depth country studies (OECD 2021a, 2022b; OECD and KDI 2018). Adjusted and improved versions of these questions have been selected by the European Social Survey (ESS) to be included in their Cross-National Online Survey (CRONOS 2) in 2021.

The measurement approach for the competence and value drivers of institutional trust moves away from perceptions and focuses instead on specific situations. Situational questions present respondents with a stereotypical situation involving the interaction of people with public institutions and inquire about its expected outcome. The deconstruction of situational questions allows for analysis of the kind of behavior under scrutiny. Typical behavioral questions, as used in psychology or sociology, investigate the subjective reaction expected from individuals in a specific situation. Complementary- and confirmatory-mechanism experiments are suggested to see whether individuals stick to their revealed choices.

However, these situational questions are not stereotypical behavioral questions: they don't focus on individual behavior but rather on the positive conduct expected of a third party: in this case, public institutions.

TABLE 28.3 Examples of Questions on the Determinants of Public Trust

Dimension	Example question
The following questions are about your expectations of behavior of public institutions. Please respond on a scale from 0 to 10, where 0 means very unlikely and 10 means very likely.	
Responsiveness	If many people complained about a public service that is working badly, how likely or unlikely do you think it is that it would be improved?
Reliability	If a serious natural disaster occurred in [country], how likely or unlikely do you think it is that existing public emergency plans would be effective in protecting the population?
Openness	If a decision affecting your community is to be made by the local government, how likely or unlikely do you think it is that you would have an opportunity to voice your views?
Integrity	If a government employee is offered a bribe in return for better or faster access to a public service, how likely or unlikely is it that they would accept it?
Fairness	If you or a member of your family would apply for a government benefit or service (e.g., unemployment benefits or other forms of income support), how likely or unlikely do you think it is that your application would be treated fairly?

Source: 2021 OECD Trust Survey (Nguyen et al. 2022).

For this reason, they measure the *trustworthiness* of a given institution or public agent. Unlike attitudes (passive response) and behaviors (active response), trustworthiness is based on the expectation of positive behavior that lies at the heart of the working definition of trust being considered. In general terms, a situational approach to measuring trustworthiness is based on the following type of question: “If a certain situation happens, how likely or unlikely is it that [public institution] will do [expected positive behavior]?” (see table 28.3).

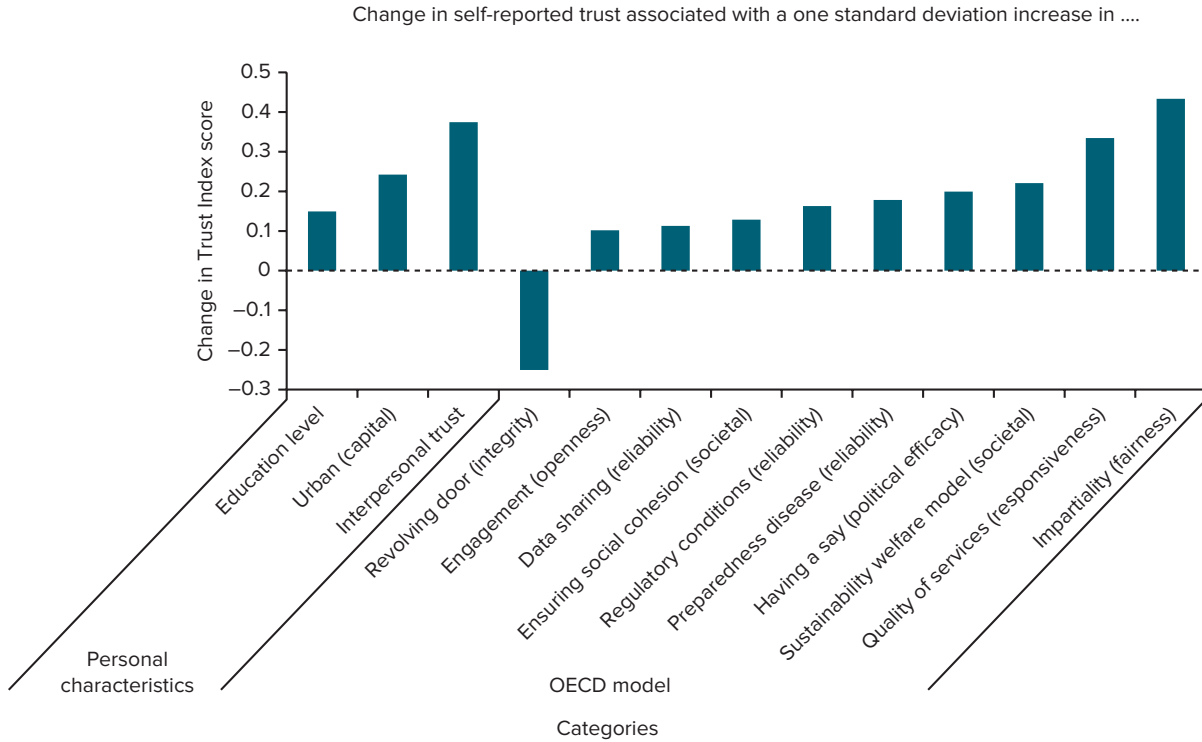
The module of the survey on the five public governance drivers of trust helps practitioners understand not only people’s perception of their government’s responsiveness, reliability, openness, fairness, and integrity but also which of these five components has a larger impact on the level of trust toward specific institutions. This evidence can provide guidance to improve public administration’s effectiveness.

SOME RESULTS FROM APPLYING THE OECD TRUST SURVEY

Data collected through the OECD Trust Survey go beyond measuring trust levels and aim at identifying what drives trust in public institutions. As an example, in the case of Norway, the OECD Trust Survey shows that the most important determinants of trust in the civil service in Norway are impartial treatment when addressing the administration and responsiveness to people’s demands for service improvement. Other societal (for example, the sustainability of the welfare model) and personal (for example, living in the capital or being more educated) characteristics also have significant, although small, relative correlations (figure 28.1).

Along the same lines, the OECD Trust Survey in Finland finds that the government’s responsiveness and reliability are the main drivers of trust, but drivers of trust vary among institutions. Specifically, in Finland, the trust relationship between people and their institutions is strongly correlated with the perceived high competence of the government and the civil service and more tenuously with values such as integrity, openness, and fairness, most likely because the latter are recognized as entrenched in public sector culture. Figure 28.2 shows the trust payoff if all significant elements pertaining to competence and values increase by one standard deviation. The responsiveness of public services and the reliability of the government in addressing future challenges and providing a stable economic environment have a greater effect on trust in the national government and civil service, while engagement opportunities are more important for explaining trust toward local governments. These data, combined with qualitative analysis and international policy dialogue, have enabled a series of policy recommendations to preserve and strengthen the trust capital in these countries (OECD 2021a, 2022b).

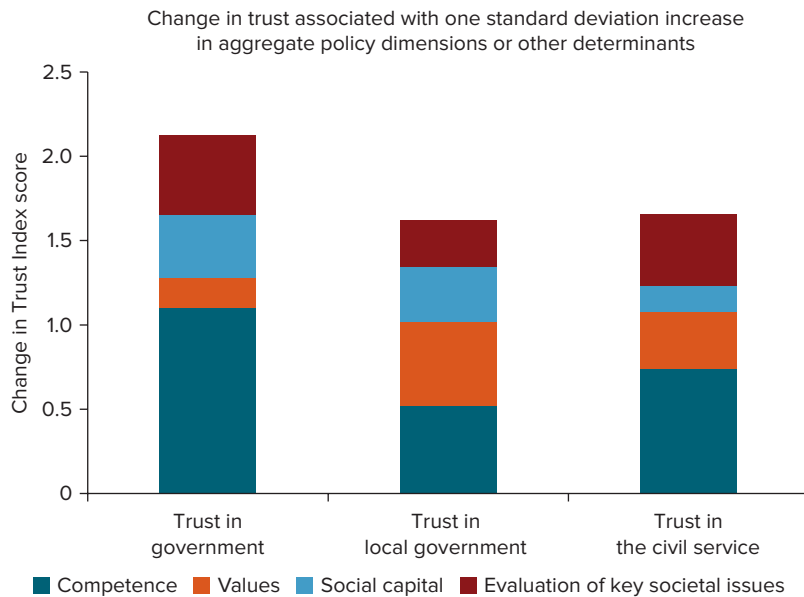
FIGURE 28.1 Determinants of Trust in the Civil Service in Norway



Source: OECD 2022b.

Note: This figure shows the most robust determinants of self-reported trust in government in an ordinary least squares (OLS) estimation that controls for individual characteristics. All variables depicted are statistically significant at 99 percent. The policy dimension is shown in parentheses. OECD = Organisation for Economic Co-operation and Development.

FIGURE 28.2 Trust Payoff Associated with Increase in Public Institutions' Competence and Values in Finland, 2020



Source: OECD 2021a.

Note: This figure shows the results of a statistical analysis to calculate the most robust determinants of trust in government. The figure shows that if competence increases by one standard deviation, then trust in the national government would increase by 1.10, trust in the local government would increase by 0.52, and trust in the civil service would increase by 0.74. If the values dimension increases by one standard deviation, then trust in the national government would increase by 0.18, trust in the local government would increase by 0.50, and trust in the civil service would increase by 0.34. Other elements, such as social capital (interpersonal trust) and the evaluations of current policies to address future challenges, are also correlated with trust in the different institutions.

Different countries have pursued different strategies to act upon these data. For example, in 2018, the Korean administration set a numerical target for its innovation strategy to improve trust levels. To achieve this target, it also included a series of actions: among others, generalizing open-government principles in different instances of the administration as a review of hiring procedures to ensure the right mix of skills for enhancing innovation within the administration. Likewise, the Finnish government has set up an inter-agency expert group to discuss concrete actions based on the data—for instance, reforming the process for formulating government policies to ensure better coordination for the inclusion of subjects such as climate change, intergenerational justice, and the preservation of social cohesion. In addition, it promotes citizenship education programs and engagement opportunities in policy choices for improving levels of political empowerment.

Based on the body of evidence developed so far, the OECD Trust Survey (based on table 28.2) has been carried out in 22 OECD countries, and the results were published in June 2022 (OECD 2022a). In addition, a number of briefs, country analyses, working papers, and country dialogues will be developed based on the OECD Trust Database.

CONCLUSION

This chapter has taken stock and presented evidence of the OECD's work in understanding the determinants of institutional trust and improving their measurement. The evidence developed so far sheds light on the feasibility and pertinence of including questions on public trust and its drivers in regular household surveys as well as their relevance for informing policy developments. While politically sensitive, there are no a priori reasons why measures of institutional trust could not be collected regularly and be subject to the same quality standards and requirements that apply to other social, economic, and environmental statistics, such as those produced by NSOs.

The OECD Trust Survey will provide international benchmarks on people's perceptions, evaluations, expectations, and experiences with the public sector and will inform the debate on how to preserve and strengthen democratic values in OECD countries and beyond. It will also allow practitioners to observe levels of trust across different institutions and drivers of trust according to different socioeconomic characteristics. In addition, it will shed light on the relative effects of the determinants of public trust in different contexts, as well as their commonalities and differences. The results will set a course in public governance areas to improve institutional trust. The performance and pertinence of new experimental modules are still to be assessed.

There is, however, room to further develop our understanding and measurement of different trust relationships. Greater attention could be paid to interagency trust or government officials' perceptions of citizens—for example, by analyzing aspects of trust across different public agencies or trust from institutions toward citizens. It may also be possible to incorporate additional aspects that influence institutional trust into the survey as ad hoc modules that respond to specific realities in each context. While these ad hoc modules might make cross-national comparisons more challenging, they would enable countries to have greater flexibility and control over designing strategies to strengthen institutional trust.

NOTES

The chapter was prepared between October and December 2021. This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries, and to the name of any territory, city, or area.

1. Results from the OECD Trust Survey are available on the OECD's website at <https://www.oecd.org/governance/trust-in-government/>.

2. Data from the Citizen Experience Survey, now known as the Survey of Trust in Australian Public Services, are available on the Australian government's website at <https://data.gov.au/data/dataset/trustsurvey>.
3. The questionnaire for the 2019 General Population Poll can be found on the website of the World Justice Project at <https://worldjusticeproject.org/our-work/research-and-data/wjp-rule-law-index-2021/2020-wjp-rule-law-index-questionnaires>.
4. Other types of validity, referred to as *discriminant validity* and *predictive validity*, are encompassed by construct validity, and the corresponding analysis is presented in the OECD Guidelines on Measuring Trust (OECD 2017a, 2017b).
5. *Distrust* refers to attitudes of insecurity, cynicism, disengagement, contempt, fear, anger, and alienation that may result in withdrawal, defiance, and support for populism. It is different from *mistrust*, which is associated with a positive attitude of critical citizenship that exercises vigilance in judging the components of the political system by being cautious, constructive, and alert. While measures of *trust in government* and the *trustworthiness* of public institutions are becoming more widely tested and used since the release of the OECD Guidelines on Measuring Trust, there are still a certain number of empirical challenges in measuring distrust at the individual level, which may require additional efforts, including, for example, enhanced sampling and focus groups.

REFERENCES

- Algan, Yann, Elizabeth Beasley, Daniel Cohen, and Marital Foucault. 2018. "The Rise of Populism and the Collapse of the Left-Right Paradigm: Lessons from the 2017 French Presidential Election." Document du travail 1805, Centre pour la Recherche Économique et ses Applications, Paris.
- Algan, Yann, Elizabeth Beasley, Daniel Cohen, and Marital Foucault. 2019. *Les origines du populisme: Enquête sur un schisme politique et social*. Paris: Seuil.
- AU and UNDP (African Union and United Nations Development Programme). 2017. *Governance, Peace and Security (GPS) Data: Stock-Taking Report 2012–15*. Addis Ababa: UNDP. https://www.undp.org/sites/g/files/zskgke326/files/migration/africa/UNDP-SHASA-GPS_web.pdf.
- Bargain, Olivier, and Ulugbek Aminjonov. 2020. "Trust and Compliance to Public Health Policies in Times of COVID-19." *Journal of Public Economics* 192: 104316. <https://doi.org/10.1016/j.jpubeco.2020.104316>.
- Blind, Peri K. 2007. "Building Trust in Government in the Twenty-First Century: Review of Literature and Emerging Issues." 7th Global Forum on Reinventing Government: Building Trust in Government, June 26–29, 2007, Vienna, Austria.
- Bouckaert, Geert. 2012. "Trust and Public Administration." *Administration* 60 (1): 91–115.
- Brezzi, Monica, Santiago González, David Nguyen, and Mariana Prats. 2021. "An Updated OECD Framework on Drivers of Trust in Public Institutions to Meet Current and Future Challenges." OECD Working Paper on Public Governance 48, OECD, Paris. <https://doi.org/10.1787/b6c5478c-en>.
- Choi, Sang Ok, and Sunhyuk Kim. 2012. *An Exploratory Model of Antecedents and Consequences of Public Trust in Government*. Unpublished paper. <https://iiustrust.files.wordpress.com/2012/07/an-exploratory-model-of-antecedents-and-consequences-of-public-trust-in-government.pdf>.
- Fairbrother, Malcolm, Gustaf Arrhenius, Krister Bykvist, and Tim Campbell. 2021. "Governing for Future Generations: How Political Trust Shapes Attitudes towards Climate and Debt Policies." *Frontiers in Political Science* 3: 656053. <https://doi.org/10.3389/fpos.2021.656053>.
- Fukuyama, Francis, and Francesca Recanatini. 2018. "Beyond Measurement: What Is Needed for Effective Governance and Anti-corruption Reforms?" In *Governance Indicators: Approaches, Progress, Promise*, edited by Helmut K. Anheier, Matthias Haber, and Mark A. Kayser, 43–70. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/oso/9780198817062.003.0003>.
- González, Santiago, Lara Fleischer, and Marco Mira d'Ercole. 2017. "Governance Statistics in OECD Countries and Beyond: What Exists, and What Would Be Required to Assess Their Quality?" OECD Statistics Working Paper 79, OECD, Paris. <https://dx.doi.org/10.1787/c0d45b5e-en>.
- González, Santiago, and Conal Smith. 2017. "The Accuracy of Measures of Institutional Trust in Household Surveys: Evidence from the OECD Trust Database." OECD Statistics Working Paper 87, OECD, Paris. <https://doi.org/10.1787/d839bd50-en>.
- Györfy, Dóra. 2013. *Institutional Trust and Economic Policy: Lessons from the History of the Euro*. Budapest: Central European University Press.
- Kaufmann, Daniel, and Aart Kraay. 2007. "Governance Indicators: Where Are We, Where Should We Be Going?" Policy Research Working Paper, World Bank, Washington, DC. <https://doi.org/10.1596/1813-9450-4370>.
- Murtin, Fabrice, Lara Fleischer, Vincent Siegerink, Arnstein Aassve, Yann Algan, Romina Boarini, Santiago González, Zsuzsanna Lonti, Gianluca Grimalda, Rafael Hortala Vallve, Soonhee Kim, David Lee, Louis Putterman, and Conal Smith. 2018. "Trust and Its Determinants: Evidence from the Trustlab Experiment." OECD Statistics Working Paper 89, OECD, Paris. <https://doi.org/10.1787/869ef2ec-en>.

- Nguyen, David, Valérie Frey, Santiago González, and Monica Brezzi. 2022. "Survey Design and Technical Documentation Supporting the 2021 OECD Survey on Drivers of Trust in Government Institutions." OECD Working Paper on Public Governance 53, OECD, Paris. <https://dx.doi.org/10.1787/6f6093c5-en>.
- Nooteboom, Bart. 2006. "Social Capital, Institutions and Trust." CentER Discussion Paper Series 2006–35, CentER Graduate School for Economics and Business, Tilburg University, Tilburg, The Netherlands. <http://dx.doi.org/10.2139/ssrn.903747>.
- OECD (Organisation for Economic Co-operation and Development). 2013. *Government at a Glance 2013*. Paris: OECD Publishing. https://doi.org/10.1787/gov_glance-2013-en.
- OECD (Organisation for Economic Co-operation and Development). 2017a. *OECD Guidelines on Measuring Trust*. Paris: OECD Publishing. <https://dx.doi.org/10.1787/9789264278219-en>.
- OECD (Organisation for Economic Co-operation and Development). 2017b. *Trust and Public Policy: How Better Governance Can Help Rebuild Public Trust*. OECD Public Governance Reviews. Paris: OECD Publishing. <http://dx.doi.org/10.1787/9789264268920-en>.
- OECD (Organisation for Economic Co-operation and Development). 2021a. *Drivers of Trust in Public Institutions in Finland*. Paris: OECD Publishing. <https://dx.doi.org/10.1787/52600c9e-en>.
- OECD (Organisation for Economic Co-operation and Development). 2021b. *Government at a Glance 2021*. Paris: OECD Publishing. <https://doi.org/10.1787/1c258f55-en>.
- OECD (Organisation for Economic Co-operation and Development). 2022a. *Building Trust to Reinforce Democracy: Main Findings from the 2021 OECD Survey on Drivers of Trust in Public Institutions*. Paris: OECD Publishing. <https://doi.org/10.1787/b407f99c-en>.
- OECD (Organisation for Economic Co-operation and Development). 2022b. *Drivers of Trust in Public Institutions in Norway*. Paris: OECD Publishing. <https://doi.org/10.1787/81b01318-en>.
- OECD and KDI (Korea Development Institute). 2018. *Understanding the Drivers of Trust in Government Institutions in Korea*. Paris: OECD Publishing. <https://dx.doi.org/10.1787/9789264308992-en>.
- Rothstein, Bo. 2013. "Corruption and Social Trust: Why the Fish Rots from the Head Down." *Social Research: An International Quarterly* 80 (4): 1009–32. <https://dx.doi.org/10.1353/sor.2013.0040>.
- Tabellini, Guido. 2008. "Institutions and Culture." *Journal of the European Economic Association* 6 (2–3): 255–94. <https://doi.org/10.1162/JEEA.2008.6.2-3.255>.
- United Nations Praia Group on Governance Statistics. 2020. *Handbook on Governance Statistics*. <https://paris21.org/news-center/news/new-praia-city-group-handbook-governance-statistics>.
- Van de Walle, Steven, and Geert Bouckaert. 2003. "Public Service Performance and Trust in Government: The Problem of Causality." *International Journal of Public Administration* 26 (8–9): 891–913. <http://dx.doi.org/10.1081/pad-120019352>.
- Van de Walle, Steven, and Koen Migchelbrink. 2020. "Institutional Quality, Corruption, and Impartiality: The Role of Process and Outcome for Citizen Trust in Public Administration in 173 European Regions." *Journal of Economic Policy Reform* 25 (1): 1–19. <http://dx.doi.org/10.1080/17487870.2020.1719103>.

CHAPTER 29

Government Analytics Using Measures of Service Delivery

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

SUMMARY

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

ANALYTICS IN PRACTICE

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

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

- Designing MSDs allows practitioners to specify the dimensions of service quality and construct measures to identify how well services perform in each one of them. Developing MSDs for primary health care and education requires considering and defining what dimensions will be used to measure the quality of public services. For example, a personnel dimension may measure absence rates for teachers and doctors.¹ Another dimension may be the availability of learning materials in schools and medical supplies in health facilities. There is a variety of conceptual frameworks and indicators to draw from. For health MSDs, practitioners may build on the existing frameworks described by *The Lancet Global Health Commission* (Hanson et al. 2022) or the World Health Organization’s “building blocks” framework (WHO 2010).
- The implementation of MSDs should follow a sequential structure, from defining a conceptual framework around measurements of the quality of public services to disseminating findings to government stakeholders and citizens. Generally, the first step in the implementation process is defining a conceptual framework and securing institutional support. The next step is identifying what will be measured: the indicators of interest; the questions to be asked; and to whom, where, and with what frequency they will be asked. After defining these indicators, practitioners should develop a rollout strategy for the actual data collection, which could include procuring a survey firm or developing a specific management information system for health care or education. After the data are collected, they should be validated, processed, and transformed into indicators. The final step is crucial: the resulting MSDs should be clearly articulated to stakeholders and disseminated widely, both within the government and to citizens.
- While practitioners may develop MSDs independently, the development of objective measurements to improve public services is part of a global agenda. Connecting to this broader movement allows practitioners to learn from other governments’ experiences. Engagement with global partners can also accelerate the design and implementation of MSDs. This global engagement can raise awareness of the relative standing of countries through benchmarking exercises, as well as facilitate knowledge exchange.
- MSDs should be subject to constant revision, as the understanding of quality in service delivery evolves. Indicators should also respond to new and unexpected demands. MSDs should evolve according to the changing policy objectives of stakeholders and citizens. Adaptations in measurement methodologies reflect an ongoing dialogue between policy makers, citizens, and the practitioners responsible for producing these indicators. As the COVID-19 (coronavirus) pandemic has highlighted, moments of crisis may generate demand for additional indicators, such as the availability of vaccines and the impact of school shutdowns on student learning.

INTRODUCTION

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

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

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

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

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

AN OVERVIEW OF MEASURES OF SERVICE DELIVERY

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

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

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

(continues on next page)

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

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

Source: Original table for this publication.

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

- Set priorities regarding improvements in service delivery.
- Identify strengths and gaps in delivery system performance.
- Identify knowledge gaps, where deeper diagnostics are needed.
- Monitor progress on the quality of services.

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

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

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

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

SDI Health Survey Conceptual Framework

The SDI Health Survey is a nationally representative, health facility–based survey that measures the quality of delivery of primary health care services as experienced by citizens across the world.² Since its inception in 2008, the objective of the SDI Health Survey has been to improve the monitoring of service delivery to increase public accountability and good governance, as well as targeted interventions through objective measurement of the quality and performance of health services. SDI Health Surveys have been completed in several countries in Africa, including Kenya, Madagascar, and Mozambique (see figure 29.1), and the survey has recently expanded to South Asia (Bhutan), Europe and Central Asia (Moldova), and the Middle East and North Africa (Iraq).

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

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

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

The Global Education Policy Dashboard Conceptual Framework

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

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



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

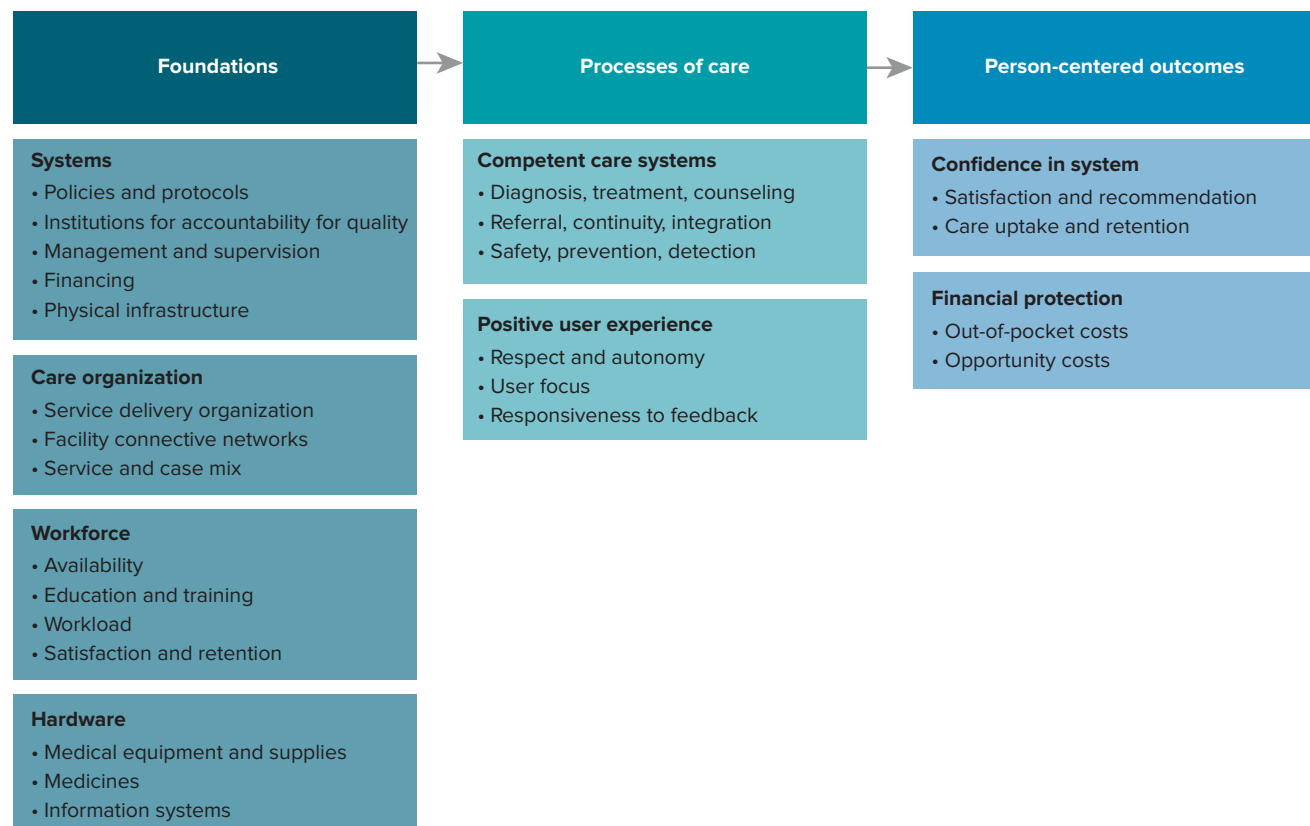
The GEPD identifies three key dimensions driving learning outcomes: practices (or service delivery), policies, and politics (figure 29.3). The practices dimension is further divided into four topics: teachers, learners, school inputs, and school management. Because the GEPD also provides indicators for learning, it offers a holistic view of the educational system, connecting outcomes (learning) to their drivers. A total of 39 indicators have been developed to measure these different dimensions. Examples of indicators, as well as their associated topics and dimensions, are provided in table 29.3.⁹

TABLE 29.2 Indicators in the First Service Delivery Indicators Health Surveys

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

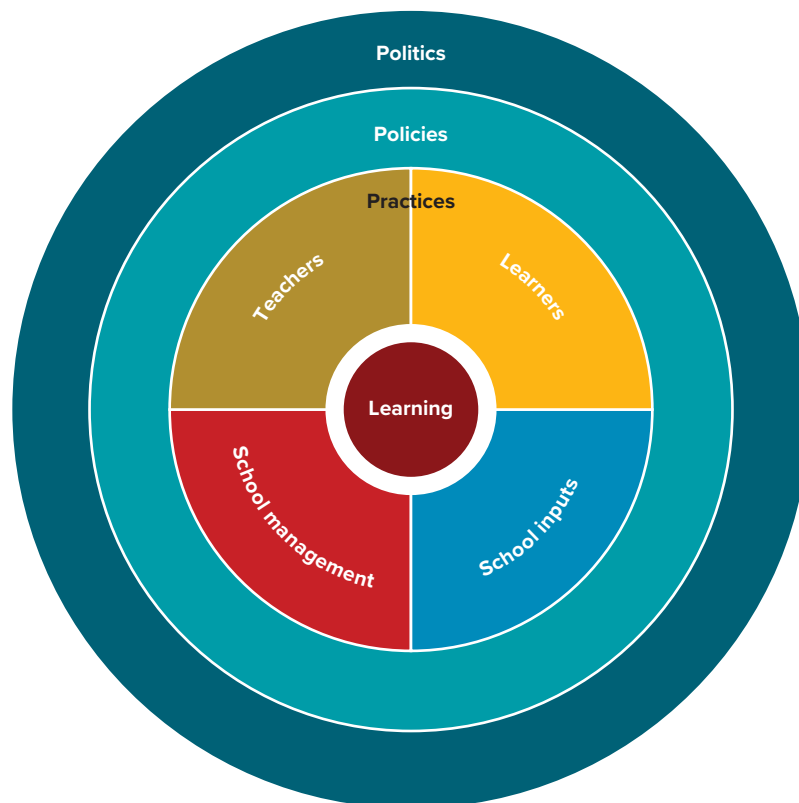
Source: Adapted from Gatti et al. 2021.

FIGURE 29.2 Updated Conceptual Framework and Questionnaire Modules for the Service Delivery Indicators Health Survey



Source: Original figure for this publication.

FIGURE 29.3 Dimensions of the Global Education Policy Dashboard



Source: Global Education Policy Dashboard website, <https://www.educationpolicydashboard.org/>.

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

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

PRODUCING MEASURES OF SERVICE DELIVERY

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

TABLE 29.3 Global Education Policy Dashboard Indicators

Dimension	Topic	Indicators
Learning	Proficiency on GEPD assessment	Each question on the fourth-grade student assessment is scored as one point. The indicator reports the fraction of students scoring at least 20 out of 24 points on the fourth-grade language assessment and at least 14 out of 17 points on the math assessment.
	Learning Poverty indicator	The Learning Poverty indicator, as reported in the Learning Poverty Database, captures schooling and learning at the end of primary school.
	Teacher content knowledge	This indicator measures the percentage of teachers scoring at least 80 percent correct on the teacher assessment. In this assessment, each question is worth one point.
	Instructional leadership	A score of one to five is assigned based on the presence of four practices as reported by teachers. The four practices, which are given equal weight, are the following: <ul style="list-style-type: none"> • Had a classroom observation in past year • Had a discussion based on that observation that lasted longer than 30 minutes • Received actionable feedback from that observation • Had a lesson plan and discussed it with another person.
Policies	Teaching—attraction	A score of one to five is assigned based on five factors. Each factor receives an equal weight in terms of possible points (0.8). The factors are the following: <ul style="list-style-type: none"> • Job satisfaction • Community satisfaction • Perceived meritocracy • Financial incentives • Absence of salary delays.
	School management—evaluation	A score of one to five is assigned based on four factors. Each factor receives an equal weight in terms of points. The factors are the following: <ul style="list-style-type: none"> • Reported evaluation in the past year (1) • Reported multiple evaluation criteria (1) • Reported consequences for negative evaluation (1) • Reported consequences for positive evaluation (1).
Politics	Characteristics of bureaucracy	A score of one to five is assigned based on four factors. Each factor has been given an equal weight. Each factor is based on a set of three to four questions, each scored one to five. For each factor, the average score across the questions is determined. To construct the total score, the average is taken of the four factor scores. The factors include the following: <ul style="list-style-type: none"> • Knowledge and skills • Work environment • Merit • Motivation.
	Impartial decision-making	A score of one to five is assigned based on four factors. Each factor has been given an equal weight. Each factor is based on a set of three questions scored one to five. For each factor, the average score across the three questions is determined. To construct the total score, the average is taken of the four factor scores. The factors include the following: <ul style="list-style-type: none"> • Politicized personnel management • Politicized policy making • Politicized policy implementation • Employee unions as facilitators.

Source: Adapted from GEPD 2021b.

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

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

Design: Stakeholder Engagement and Survey Instrument

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

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

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

Implementation: Training and Data Collection

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

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

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

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

Analysis: Data Validation and Production of Indicators

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

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

Dissemination: Stakeholder Presentations and Wider Communication Efforts

The final step in the production cycle for MSDs is the dissemination of the results to stakeholders, as well as wider communication efforts. In general, we have found that interactive dashboards are an important component of MSD dissemination. Dashboards allow for intuitive visualization of the different dimensions of public service and empower stakeholders to interact with the data at a greater level of granularity than static reports. For an example, see figure 29.4.¹⁷ Practitioners may click on different dimensions and obtain additional information for particular indicators. Colors allow users to identify where in the public service delivery chain more attention is needed.

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

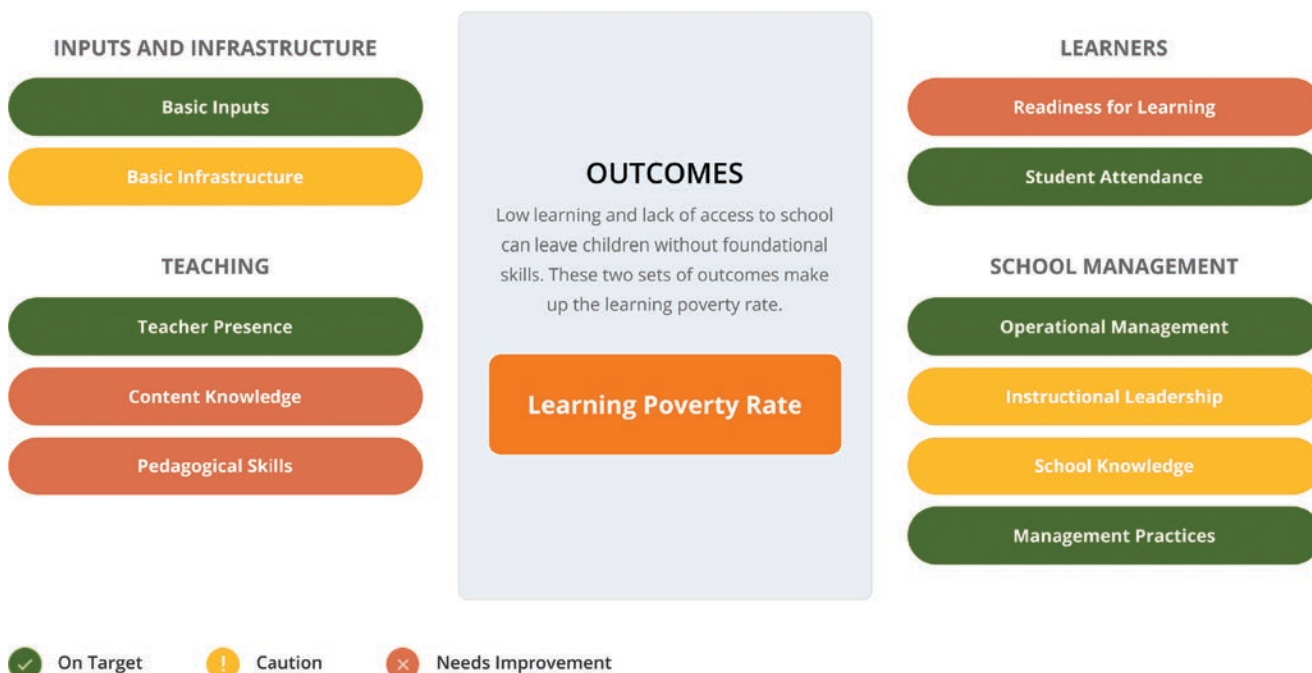
This systematic view, with diagnostics associated with each topic, also allows practitioners to hone in on specific areas that require further development. For instance, in figure 29.4, we find that Peru has done quite

FIGURE 29.4 Example of Global Education Policy Dashboard

Peru



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



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

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

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

A GLOBAL EFFORT TO IMPROVE PUBLIC SERVICES

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

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

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

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

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

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

- Standardized health facility survey materials, field manuals, training materials, and suggestions for adaptation of the survey instrument
- Technical guidance on survey design and sampling strategy
- Assistance with quality control during data collection
- Capacity building for governments to generate and disseminate MSDs.

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

CONCLUSION

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

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

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

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

NOTES

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

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

REFERENCES

- Amin, Samia, Jishnu Das, and Markus Goldstein, eds. 2007. *Are You Being Served? New Tools for Measuring Service Delivery*. Washington, DC: World Bank.
- Andrews, Kathryn, Ruben Conner, Roberta Gatti, and Jigyasa Sharma. 2021. "The Realities of Primary Care: Variation in Quality of Care across Nine Countries in Sub-Saharan Africa." Policy Research Working Paper 9607, World Bank, Washington, DC.
- Andrews, Kathryn, and Jigyasa Sharma. 2021. "A Revolution in Health Service Delivery Measurement." *Investing in Health* (blog). *World Bank Blogs*, July 1, 2021. <https://blogs.worldbank.org/health/revolution-health-service-delivery-measurement>.
- Björkefur, Kristoffer, Luíza Cardoso de Andrade, Benjamin Daniels, and Maria Ruth Jones. 2021. *Development Research in Practice: The DIME Analytics Data Handbook*. Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/35594>.
- Gatti, Roberta, Kathryn Andrews, Ciro Avitabile, Ruben Conner, Jigyasa Sharma, and Andres Yi Chang. 2021. *The Quality of Health and Education Systems across Africa: Evidence from a Decade of Service Delivery Indicators Surveys*. Washington, DC: World Bank.
- GEPD (Global Education Policy Dashboard). 2019. "School Data Quality Checks." School, Master Code, GEPD, World Bank. GitHub repository, September 24, 2019. https://github.com/worldbank/GEPD/blob/master/Master_Code/School/school_data_quality_checks.Rmd.
- GEPD (Global Education Policy Dashboard). 2021a. "Global Education Policy Dashboard Implementation Brief." Global Education Policy Dashboard, World Bank Group. <https://www.educationpolicydashboard.org/sites/epd/files/resources-documents/GEPD%20Implementation%20Brief.pdf>.
- GEPD (Global Education Policy Dashboard). 2021b. "GEPD Technical Note." Global Education Policy Dashboard, World Bank Group, April 19, 2021. <https://www.educationpolicydashboard.org/sites/epd/files/resources-documents/GEPD%20Technical%20Note.pdf>.
- Hanson, Kara, Nouria Brikci, Darius Erlangga, Abebe Alebachew, Manuela De Allegri, Dina Balabanova, Mark Blecher, Cheryl Cashin, et al. 2022. "The Lancet Global Health Commission on Financing Primary Health Care: Putting People at the Centre." *The Lancet Global Health Commission* 10 (5): E715–E772. [https://doi.org/10.1016/S2214-109X\(22\)00005-5](https://doi.org/10.1016/S2214-109X(22)00005-5).
- Kruk, Margaret E., Anna D. Gage, Catherine Arseneault, Keely Jordan, Hannah H. Leslie, Sanam Roder-DeWan, Olusoji Adeyi, et al. 2018. "High-Quality Health Systems in the Sustainable Development Goals Era: Time for a Revolution." *The Lancet Global Health Commission* 6 (11): E1196–E1252. [https://doi.org/10.1016/S2214-109X\(18\)30386-3](https://doi.org/10.1016/S2214-109X(18)30386-3).
- SDI (Service Delivery Indicators). 2019. "Nuts and Bolts: A Brief Guide for Task Teams." Unpublished document, September 24, 2019, World Bank, Washington, DC.
- WHO (World Health Organization). 2010. *Monitoring the Building Blocks of Health Systems: A Handbook of Indicators and Their Measurement Strategies*. Geneva: WHO Press. <https://apps.who.int/iris/bitstream/handle/10665/258734/9789241564052-eng.pdf>.
- World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank. <https://www.worldbank.org/en/publication/wdr2018>.
- World Bank. 2020. "Systems Approach for Better Education Results (SABER)." Education Brief, World Bank. <https://www.worldbank.org/en/topic/education/brief/systems-approach-for-better-education-results-saber>.
- World Bank. 2021. "What Is Learning Poverty?" Education Brief, World Bank, April 28, 2021. <https://www.worldbank.org/en/topic/education/brief/what-is-learning-poverty>.

CHAPTER 30

Government Analytics Using Anthropological Methods

*Colin Hoag, Josiah Heyman, Kristin Asdal, Hilde Reinertsen,
and Matthew Hull*

SUMMARY

This chapter aims to present an overview of how anthropologists study bureaucracy and why that approach has value to the World Bank and its interlocutors. Anthropologists are most commonly associated with immersive ethnographic methods, such as participatory observation. In this chapter, we describe those methods and their usefulness, but we also highlight the heterogeneity of the empirical materials that anthropologists draw upon. The chapter makes the case that, while the ethnographic approach of anthropologists might sometimes be perceived as “messy” or “unstructured,” in fact, the efforts of anthropologists are motivated by an abiding concern with empirical rigor—a refusal to ignore any sort of data or to content oneself with a single view of such a multifarious thing as bureaucracy. This is to say that an anthropological approach is a holistic one, which envisions bureaucracy as a rich, multidimensional world.

ANALYTICS IN PRACTICE

- Study the bureaucratic process or organization you are interested in holistically by observing all aspects of it: engage with the staff who are involved at every level of the organization, from senior officers to low-level staff and contractors, and with different demographic groups; study everyday documents; and watch how officials interact. Observe every part of what they do at work holistically, from their interactions in corridors and meetings to the protocols they observe in their relationships. Processes or organizational outcomes may be shaped by forces outside of those that the analyst presupposes.
- Develop relationships with a variety of people and have open-ended conversations about their work as well as about unrelated issues to understand their values and perspectives.

Colin Hoag is an assistant professor at Smith College. Josiah Heyman is a professor at the University of Texas at El Paso. Kristin Asdal is a professor and head of the Centre for Technology, Innovation and Culture at the University of Oslo. Hilde Reinertsen is a researcher at the Centre for Technology, Innovation and Culture at the University of Oslo. Matthew Hull is an associate professor and director of the Center for South Asian Studies at the University of Michigan.

- Engage in participatory observation by observing government activities in person as they unfold. This practice can capture activities that may be so routine that they go unnoticed by public officials and are not self-reported in surveys. Practitioners should talk with those involved in government processes, both public officials and clients, as they participate in them. When attending meetings, practitioners should examine them as a form of social engagement where formal and informal rules preside and the process of forming consensus is revealed. Practitioners should likewise examine not only the content of documents but how they are developed and circulated and how they seek to further broader policy goals.
- Aim to collect the widest practical range and amount of data, both qualitative and quantitative, even if it cannot be easily standardized. Data may be gleaned from studying aspects of everyday life: the way documents are read; meetings are run; and policies, media, and whatever else are perceived as relevant. There is a trade-off in all analyses between standardization and holistic measurement. Ensure a role for holistic measurement in your approach.
- Study the heuristics, interactions, scope for discretion, and microscopic decisions that affect the gap between stated policy goals and the actual work being carried out by public officials. Real-world manifestations of policy depend on the decisions of individual officials about how to interpret broader policy goals. Using the words and actions of officials, practitioners should aim to determine which factors contribute to how officials make decisions and prioritize tasks. In this process, consider that these decisions may result from learned behaviors and rationales that require extensive effort to change.
- Revise research questions and the focus of analysis as the study progresses. As observation and insight reveal new lines of inquiry, practitioners should be open to shifting their questions and methods. The initial research plan (including the research questions and methods) should be considered provisional.

INTRODUCTION

This chapter aims to present an overview of how anthropologists study bureaucracy and why this approach has value to the World Bank and its interlocutors. Anthropologists are most commonly associated with immersive ethnographic methods, such as participatory observation. In the pages that follow, we describe those methods and their usefulness, but we also highlight the heterogeneity of the empirical materials that anthropologists draw upon.

An anthropological analysis might include an ethnographic rendering of the sights, sounds, spatial organization, and everyday life of a bureaucratic workplace. But it might also include a discourse analysis of policies that emanate from the federal government or “head office,” including the categories of persons and practices that those policies define, as well as a consideration of how such policies articulate with office life and bureaucrats’ professional fears, aspirations, and values. It could include an analysis of documents and other artifacts produced through bureaucratic labor and how these documents move within and beyond political and bureaucratic sites, or of the social interactions that give shape to office life, such as meetings, performance reviews, and dealings with clients. The world outside of the bureaucracy—after-work gatherings, access brokers hanging around the main doors of the office, or an individual bureaucrat’s living situation—might be as relevant to an anthropologist as the world within the office walls.

Anthropologists do not merely study bureaucratic activities at the “street level”; they examine practices from the bottom to the top of organizational hierarchies, as well as those who engage organizations from the outside. Anthropological work is especially attuned to the informal aspects of bureaucratic practices, but it also examines the surprising aspects and effects of formal policies and procedures.¹ This chapter makes the case that, while the ethnographic approach of anthropologists might sometimes be perceived as “messy” or “unstructured,” in fact, the efforts of anthropologists are motivated by an abiding concern with empirical rigor—a refusal to ignore any sort of data or to content oneself with a single view of such a multifarious thing

as bureaucracy. This is to say that an anthropological approach is a holistic one, which envisions bureaucracy as a rich, multidimensional world.

The chapter is organized around four topics that capture this multidimensional approach, and which have been of particular focus to anthropologists. The first section, by Colin Hoag, covers the *everyday life* of bureaucratic institutions, showing why attention to informal practices and the flow of office life can be critical to understanding the workings of formal organizations. Second, Kristin Asdal and Hilde Reinertsen discuss bureaucratic documents, followed by Colin Hoag and Matthew Hull describing the significance of meetings to bureaucratic life. Finally, Josiah Heyman describes the anthropological approach to policy. Throughout, the aim is not to offer an encyclopedic account or to describe the state of the field but rather to show how anthropologists approach bureaucracy and why such an approach has value to policy makers. We hope to show why anthropological sensibilities might offer a meaningful guide for World Bank policy making and for crafting goals and guidelines that are informed by culture, power, and everyday life.

EVERYDAY LIFE

Anthropology's early interest in the exotic and non-Western has meant that anthropologists have been relatively quiet on the topic of bureaucracy when compared with sociologists and political scientists. This has changed dramatically in recent years, partly out of dissatisfaction with accounts of powerful institutions from the outside. Commenting on research about the state in Africa, for example, the anthropologist Jean-Pierre Olivier de Sardan (2009, 39) explains, "It is easy to get the feeling that, for decades, journalists, politicians and many researchers, both Africans and Africanists, have been engaged in a relentless search for the 'essence' of the African state while neglecting to carry out a concrete analysis of the administrations, public services, bureaucratic system and relations between civil servants and the users of state services." His point is emblematic of a broader commitment among anthropologists of bureaucratic institutions to develop an empirical record of how such organizations function. Their approach seeks to understand bureaucracies as rich lifeworlds rather than organizational charts and protocols. Using anthropological methods of participant observation to understand how bureaucracies work leads to a fundamentally different picture of bureaucracy than is conventionally given by political scientists and public administrators.

A critical concept for anthropologists in general, and certainly for anthropologists of bureaucracy, is *everyday life*. The term refers to all those routine or unremarkable things that are so common they might even go unnoticed by bureaucrats, but which constitute the bulk of bureaucratic activity. Anthropology may have a reputation for caring about the exotic and about major cultural events, such as religious rituals. In fact, many anthropologists focus on the routine and quotidian aspects of culture. In the case of the civil service, anthropologists are likely to be less interested in the pronouncements of top administrators than the flow of ordinary events at the office.

These anthropologists could be said to work in the tradition of scholars of public administration and organizational ethnography from the early 20th century, such as Chester Barnard. Those scholars sought to explain the role of informal practices to the functioning of formal rules. What factors determine why bureaucracies work in the ways they do? Do bureaucrats determine institutional practice, or are they controlled by institutional rules and regulations? How do the effects of bureaucracies correspond to their stated aims? What factors other than formal rules and regulations (for example, institutional history or culture) influence bureaucrats? How do bureaucrats interpret or experience their work—the rules and reforms that guide them; the clients, bosses, or employees with whom they interact; and their own actions?

Accounting for this form of everyday life requires an immersive approach. Anthropologists employ participant observation, a method that entails joining bureaucrats in their work and recording what bureaucrats do, as opposed to simply relying on their verbal or written responses to questions on a survey or interview protocol. This shift of focus helps to remedy well-established problems associated with biased self-reporting.

Instead of recording what bureaucrats *say* they do, anthropologists attempt to record what they *actually* do. Participant observation can appear haphazard and time-consuming, and it indeed entails a substantial amount of simply “hanging around” the office. However, the method opens up an experiential understanding of bureaucracy while also allowing researchers to build rapport with the people they study.

Anthropologists seek to immerse themselves in bureaucratic worlds, aiming to understand what it feels like to do bureaucratic work, based on the assumption that this *feel*—this embodied sense of office life—shapes how the formal rules governing an organization take shape. Are the sights and sounds of an office taxing? For a bureaucrat making repetitive but consequential decisions, this might lead to inconsistency (or, by contrast, a lack of attention and nuance to a given decision). Such an approach also highlights the heterogeneity of office life, rather than presuming that *organizational culture* encompasses all social life. For example, bureaucrats’ perceptions of different spaces within an institution might differ across lines of race, class, or gender.

Anthropological approaches through participant observation strive to provide this kind of texture to account for the institution as a heterogeneous place. An employee’s career trajectory or dissatisfaction with their pay could determine how they appreciate a given policy reform. One employee interviewed during research at the Department of Home Affairs in South Africa, for example, was particularly sour after having been relegated to a position she disliked for a full year. When asked about how she viewed a departmentwide reform initiative called the “turnaround strategy,” the official replied, “But why aren’t they turning around our salaries?” In short, a policy reform that makes perfect sense in strictly public administration terms might fail if it does not account for these everyday factors.

Anthropological approaches might also extend outside the organization to conceive of how office life “spills over” into other social spaces, such as happy hours or holiday parties. By contrast, they might also be interested in describing how outside events, such as party politics, kinship structures, or even football allegiances, shape the lived experience of working at the institution. At the South African Department of Home Affairs (see Hoag 2014), the architecture and materiality of office spaces were important factors that shaped how clients accessed government services. Though rarely (if ever) referenced in government reports about reforms and service delivery, client perceptions of the opacity of the government bureaucracy owed at least partly to the fact that office counters were literally opaque: covered with taped signs and notifications. The loud din of the waiting room made it hard for them to hear the requests of the bureaucrat, and this was a source of frustration for bureaucrats. Instructions for applicants were poorly documented on department websites, meaning that applicants often required multiple trips to the office. Some solicited the support of *agents* or *immigration practitioners*, private citizens who specialized in advising applicants about the process. Bureaucrats had developed relationships with these agents, and those relationships sometimes led bureaucrats to treat applicants preferentially—for example, by excusing mistakes.

DOCUMENTS

Curiously, among the often-overlooked parts of everyday office life are the materials with which bureaucrats work: bureaucratic *documents*.² Indeed, documents are everywhere, but despite their significance, we often speak of them in negative ways: as dusty and dull, piling up on a desk, put on a shelf, or placed in a drawer. Documents tend to symbolize inefficiency, inertia, and pointless bureaucracy. The word *paperwork* itself implies something that stands in contrast to real, hands-on, meaningful work. Yet if we dismiss documents in this way, we risk overlooking their fundamental importance. Not only are documents critical in our individual lives, paperwork is itself a defining feature of modern institutions (Asdal and Reinertsen 2022). Documents, be they physical or digital, are integral to organizational practices, shape organizational culture, and thoroughly shape and reshape our societies.³

Sometimes, documents are part of deep controversies. Just think of the reports from the United Nations Intergovernmental Panel on Climate Change (IPCC). These comprehensive reports, their production

process, and their reception and use are subject to intense public and political debate. But documents are also crucial in producing trust, consensus, and agreement. In fact, documents are often also, quite literally, agreements. Other documents, such as governmental budgets, have a perhaps less visible public role but are no less influential. A government budget organizes the political year and determines public spending on roads, schools, hospitals, and all other sectors in a given country. This attests to how documents take part in shaping society. We therefore benefit much from studying documents both *in practice* and *as practice*.

Practice-based document analysis (Asdal and Reinertsen 2022) has been developed precisely to capture the significance of documents, both in organizational settings and in society more broadly. This approach delineates six methodological moves for studying documents: document *sites*, document *tools*, document *work*, document *texts*, document *issues*, and document *movements*. These six methodological moves are also simultaneously analytical concepts. In adopting a practice-oriented approach, we see that intense power struggles are in fact taking place in and around documents: Who is allowed to write? Who is the document's sender? Who and what is mentioned in the document, and who and what is not? Who is the recipient, and who is allowed to handle the document? When is it important *not* to write, report, and make a document? Individuals, groups of actors, and issue components may be defined in and out of documents and the issues they concern and shape. Documents are sources of power; they provide opportunities and spaces of action. What is happening in and behind the documents? A practice-oriented document analysis aims at exploring these kinds of questions.

In the following, we will go through these six methodological moves and show how they compose a cohesive analytical framework. But just to make the point clear: even if these elements together make up a whole, this does not mean that they are the only elements or that they are always equally relevant. This depends on the object of analysis and on what one is interested in exploring and analyzing in a specific organizational setting. The different elements are partly overlapping, and they “speak” to one another. When we now go through them in sequence, we will do so by means of the recent case of the COVID-19 (coronavirus) pandemic, to illustrate the many dimensions of documents this method allows us to explore.

As these pages were being written, governments across the globe were struggling to contain the COVID-19 pandemic. Continuously updated risk analyses, swiftly prepared emergency laws, and rapidly changing travel restrictions were but three of the many forms of documents deployed in the effort. The latter two were designed to manage us, as individual citizens, to ensure that we acted in a manner that helped contain rather than spread the virus. Yet newspapers also reported that fake negative COVID-19 test scores could be bought online, enabling individuals to escape quarantine restrictions (and risk being accused of document fraud). These are but a few examples of the many documents involved in the ongoing tracking and handling of the coronavirus. So how might we go about analyzing the coronavirus and the COVID-19 pandemic by way of documents? Indeed, if we start looking at how the pandemic unfolded in practice, we will soon see that documents were involved at every turn. In the following subsections, we analyze the coronavirus and the COVID-19 pandemic by way of the six methodological moves of the practice-oriented method (Asdal and Reinertsen 2022).⁴ In short, this will enable us to analyze and demonstrate how documents made the virus governable.

Document Sites

Documents always exist at specific sites—such as archives, websites, organizations, and bureaucratic offices—and it matters what kinds of sites these are. To understand how bureaucratic institutions have handled the COVID-19 pandemic, we can study what happens inside government offices. Yet we may also extend this site-oriented move to the documents as such. We can consider documents as sites in themselves: sites where medical facts and political decisions are formulated, negotiated, and decided upon. They are sites to which we may go, analytically speaking, to study the pandemic. This means to ask not only “What does this document tell us?” and “What is written here?” but also “What does this document do?” (that is, “What effects or force does it have?”), “What happens here?” and “What are the practices unfolding here?” In so doing, we can, in fact, think of document analysis as a form of fieldwork—a form of document ethnography.

Document Tools

In a dramatic political situation, such as an unfolding pandemic, what becomes blatantly clear is that documents are tools: they are produced and used for specific reasons, and they are part of larger processes, cases, and institutions. They are written and printed and distributed with the intention that something can and should happen by means of them. Just consider the following three documents: maps displaying countries and regions as green, yellow, and red to signify which travel restrictions are in place; emergency laws equipping the government with extra measures to act in times of crisis; and economic stimulus packages undergoing hard negotiations in the legislature before channeling government funds to industries and public services across the country. Color-coded maps determine who may travel where, emergency laws enable the government to act more independently from the legislature, and stimulus packages help actors and institutions endure a dire economic situation. When we start thinking about documents as tools, we become “tuned in” to investigating what role documents play in a specific situation and how their particular properties affect how they are used and how they shape bureaucratic outcomes—and by extension, societal outcomes.

Document Work

No document miraculously emerges in its finalized form. Producing and handling documents are, in themselves, forms of labor, craft, and expertise. Documents are always part of specific work practices, including writing, editing, circulating, reading, and use. We can study the various ways in which this work is done by getting close to the people working on and with them. Examples of such document work are the preparation of weekly governmental COVID-19 contamination reports and the updating of public guidelines online. This is often a matter of collective work within larger institutions and bureaucracies. Even though the public faces of a government’s pandemic response are high-level politicians and agency leaders, a host of staffers have been involved across ministries, directorates, and agencies. Their concerted (albeit sometimes conflictual) document work, in which all are involved in drafting separate paragraphs, reports, and reviews, is what, in combination, enabled the full COVID-19 response citizens witnessed through the media.

Document Texts

Clearly, it is essential to analyze documents as texts. This is the content, the very material, that document work concerns itself with. Paying attention to the document as text includes analyzing its genre—the textual, rhetorical, narrative, and visual properties that together make up the document in front of us. Think, for example, of the guidelines for quarantine and isolation that everyone was obliged to follow, and which have been critiqued for being hard to understand and easy to misinterpret. What rhetorical situation do the guidelines establish? How do they try to explain their topic and convince their readers? In analyzing guidelines as texts, we can look at their author, intended recipient, style, structure, layout, illustrations, graphs, numbers, and references. What are the combined effects of these elements? How do they seek to produce authority and trust? Why and how did they succeed, or not?

Document Issues

Documents are sources of information about the specific issues in which we are interested, shaping how issues are understood and acted upon. To understand what the COVID-19 pandemic is about, we retrieve documents from the government, from researchers, and from the media. Yet documents also take an active part in forming the issue itself—as, for instance, a situation that is *under control* or *out of control*, as a *global* issue or a *national* question, as an issue that is *closed* or one which is *uncertain* and open for discussion. They act upon the issue and thus have a transformative capacity that we as analysts should not only acknowledge but actively investigate. In a special case such as the COVID-19 pandemic, this potential for intervention and

transformation is readily visible. Yet in less tangible and acute issues as well, such as the implementation of environmental regulations or nature conservation, documents are key to how issues are rendered governable and regulated.

Document Movements

Documents are seldom lying still. They are often “on the move,” circulating within and across sites. In the case of the COVID-19 situation, this is true to the extent that new rules, regulations, stimulus packages, and the discussion thereof have moved throughout government bureaucracies ceaselessly. Moreover, the virus itself is made manageable by how it enters into documents; thus, this move into documents is what makes it accessible and “workable.” Documents build upon one another, enabling the pandemic issue to move through the government into the public—and often back again, for new iterations of research and regulation. Furthermore, patterns of document movement reflect and even constitute the effective organization of bureaucratic institutions.

Both alone and in combination, these six methodological moves make it possible to analyze documents as valuable sources for understanding the workings of bureaucratic organizations and beyond. In the case of the COVID-19 pandemic, they may help open up the material to rich analyses by getting closer to the role of bureaucratic institutions in the situation. In drawing viruses and documents together, we may thus understand them both better. This is true for any issue in which documents are involved: by better understanding how documents operate and the effects they cause within and beyond bureaucracies, we are also better equipped to notice and appreciate their significance.

MEETINGS

Meetings are crucial sites of bureaucratic social activity. Indeed, the projects of an organization are often constituted as systems of meetings (Brown and Green 2017). However, the mundane quality of meetings tends to obscure their cultural content. Meetings have been idealized as “the locus and embodiment of ideas of appropriate, transparent decision-making” (Brown, Reed, and Yarrow 2017, p.11), but anthropologists of bureaucracy approach meetings as cultural phenomena, with particular social rules, ritual qualities, spatial and temporal framings, and consequences (Schwartzman 1989).

While analysis of meetings might initially focus attention on the content that is decided at meetings, such as the consequences or the political content of the discussion, anthropologists also pay attention to the meeting itself as a form of social engagement. As a kind of event that requires the face-to-face presence of people, meetings are a form of social life difficult to study without ethnographic work. Anthropologists recognize that organizations are unstable, in spite of their projected coherence. Employees aspire toward promotions and come and go from the organization, and their roles are often contested. An organization’s (or unit’s) goals may be inconsistent or understood differently across the hierarchy, and technologies consistently reroute the ways that people report their work or interact socially with one another. Meetings are a key site where those ambiguities are stabilized and consensus is established or enforced, even if only temporarily.

In an analysis of National Science Foundation grant application reviews, for example, meetings serve as important sites for the collective socialization of reviewers regarding how to assess applications against guidelines (Brenneis 1999). That is, they serve to guide bureaucrats in what they should take note of and value. Meetings can also be a means to generate a shared understanding of which phenomena bureaucrats should overlook, as when meetings of Mexican environmental regulators produce official ignorance of burning and firewood cutting (Mathews 2011). Similarly, in European Union policy meetings, bureaucrats succeed in crafting policy from different national member states’ interests by withholding and strategically concealing the political content of meetings (Thedvall 2013).

Because of the regularity of meetings, which often occur in specific locations at regular intervals and with predictable formats and ways of speaking, they should be understood as having a ritual quality that anthropology is well positioned to interpret. In some contexts, the performances of authority and agreement are more important than efforts to present information and establish common understanding. For example, at meetings in Lesotho between conservation bureaucrats and the rural people who were targets of their conservation efforts, bureaucrats' performance of authority and the public's performance of consent to regulation were as important as the actual transfer of information about a policy (Hoag 2022). Organizational meetings typically combine formal and informal aspects of organizations, opening in some moments to social conflict while containing it within the framework of normal business activity. Meetings often conclude with a mechanism for, if not the reconciliation of, conflict, the redirection of conflict into activities that might resolve it—often further meetings.

POLICY

Policy is a governing statement by an organization, from a nation-state to a local clinic, shaping its actions in its environment, both social and biophysical.⁵ Yet often public statements are abstract and ideal, far from actual conduct on the ground. Blame for this gap partly rests on the organization that enacts the policy. The bureaucracy fairly or unfairly takes the blame for any failure to fulfill official aspirations. Sometimes this finger-pointing holds some truth, but bureaucratic failure and limitations are by no means the only causes, and even when bureaucracy does play a role, this only raises questions of why and how it does so, and whether it might change.

It is important not to reduce issues of administration simply to questions of identifying and applying objective scientific management. While skillful management does matter, even the choice to manage for some priorities and not others is inherently a political decision. In Mexico, for example, COVID-19 vaccination has been provided first—and with great publicity—in small, rural, and often indigenous sites. This was in some ways a suboptimal decision—mass vaccination in large cities would have reached more people, more quickly, and at a lower cost. But it was explicitly a political decision to show concern for people and places that historically have been devalued and stigmatized, places that in the past, Mexican bureaucrats would have reached slowly, if at all. There is no unambiguous, scientifically right way for bureaucracies to choose their means and ends; decisions should always be understood as political.

Without delving deeply into the study of policy whys and wherefores, a few basic observations help. Even when overly ambitious or misleading, policy statements are performative and need to be understood as such. Such statements present an ideal to diverse publics (including the bureaucracy) about what should be, with the proviso that they may be taken in a wide range of ways. They may be accepted as legitimate aspirations, even if incompletely achieved. But often there is a glaring gap between formal statements of policy and the real thrusts of organizational action, covered by rhetorical adherence to the formalities of official policy.

In observations of the work of United States immigration and border officers at the US-Mexico border, officers consistently said their overriding policy goal was to interdict terrorists. But in their actual work, they never encountered potential terrorism, whereas their operations were aimed clearly at Latin American labor migrants and asylum seekers. The verbal performance of protecting the erstwhile vulnerable homeland against terror, a hard-to-question goal, justified morally debated and ambiguous duties in reality. Policy formalisms, then, have a complex relationship to actual activities on the ground.

While policy, in one regard, is those means and goals that are publicly announced, policy can also be considered as choices shaping what actually is carried out. What if the real-world implications of bureaucratic assignments and practice constitute well-understood and tacitly accepted policy that deviates from formal statements? The just-cited example of US policy at the US-Mexico border is a clear instance. This dual view of policy as rhetoric and policy as enacted requires observing the actual work routines and the

accumulated choices—to do and, importantly, not to do—of bureaucrats in performance, not taking those features as failures but as social facts in themselves. To explore these phenomena, the notion of *street-level bureaucracy* is helpful. Street-level bureaucrats are officials who interact with the public directly or carry out the actual activities of the organization. Beyond the strict definition of *street-level*, we need also to consider lower levels of management in the organization who work in close proximity to such officers. Policy, whatever is declared, always passes through the hands of street-level bureaucrats and is enacted according to their ideas and actions.

Policies, rules, and the like are inevitably written in general terms. It would be hopeless to write them to cover all possible people and situations. Instead, officials must exercise discretion in when and how to apply actions. One person might be arrested, another ignored; one person might be awarded a valued document, another denied. Discretion is not just the arbitrary impulse of willful bureaucrats. It has specific political, social, and cultural features. For example, officials who award US temporary visiting visas to Mexicans at the US-Mexico border—a desired asset for shopping and family reunions—have guidelines that aim to prevent visa misuse for unauthorized residence or work. But general guidelines need to be applied to the circumstances of diverse applicants. Many different factors enter into judgment—for example, age, with older people deemed more trustworthy. A particularly important factor of discretionary favor is wealth because people with wealth are thought to be less likely to want to cross over from Mexico to work in the United States and are generally looked on as more worthy. They dress cleanly and neatly, present themselves in polite but relaxed ways, and often speak English. That is, they resemble US officers, even outclassing them. The idea that a wealthy person does not seek unauthorized residence or employment is often correct, but by no means always; there is a notable population of wealthy unauthorized residents inside the United States, especially in border communities. Discretion, then, deserves attention as a crucial leverage point in how policy is rendered into reality by a massive aggregation of rules of thumb, interactions, discretion, and microscopic decisions.

Woven into discretionary decisions, street-level bureaucrats often ration their efforts and outcomes, whether positive or negative. They rarely have sufficient resources (personnel, time, equipment, sites, goods and funds to distribute, and so on) to enact the entire policy in all cases, either the policy as formally stated or as tacitly understood. Rather, they prioritize action and (by implication) inaction. When all individual actions are aggregated, the allocation of rewards and punishments constitutes a *de facto* policy. The bases of this allocation are complex but discernible with close attention to the words and actions of officials. Often encountered are criteria of socially interpreted personal or moral worthiness and belonging to insider versus outsider groups. Also relevant, of course, is the cost of various actions to bureaucrats and organizations in terms of scarce time, resources, and so forth.

Bureaucrats do not act in such ways alone. Rather, they interact with a wide variety of counterparties: upper executives, political bosses, publics of varying degrees of influence and respect, other organizations, rivals or collaborators, biophysical organisms, processes and flows, and so forth. Alberto Arce and Norman Long (1993) observe in rural Mexico, revealingly, systematic maneuvers and misinterpretations in the interactions between an idealistic development project engineer and agropastoralists grounded in a local tradition of defiance, mistrust, and subterfuge. A negotiated appearance of collaboration on all sides finally breaks down in project failure. This actually reinforces the interpretations on both sides perfectly. Agency managers view the breakdown as evidence that the engineer's proposed peasant-oriented innovations were wrong all along, and peasant leaders view it as another instance of incompetence and failure by central authorities.

There are wider lessons in this specific case. First, understanding bureaucrats means not only learning about their internal ideas and routines but also seeing them as diverse toolkits for interacting with counterparts, from clients through outside visitors to funders and political bosses. Each kind of interaction, when encountered, provides only partial information about an organization, since the same bureaucrat might deal very differently with another kind of counterparty. This applies, not least, to outside development experts. Second, these webs of relations are suffused with power dynamics: some clients may be deferential supplicants, others well-wired operators with higher status and better connections than bureaucrats themselves. Policy on the ground, then, must be interpreted and enacted in this diverse and unequal web of relationships.

Interpretive judgments of people, situations, and action or inaction require human intelligence. Bureaucrats might seem rigid or unresponsive, and often they are, but they rely on learned thought routines that have worked adequately for current or past sociopolitical fields. Routines only change when those fields meaningfully alter and new work patterns are available to be learned. A well-socialized official also learns to justify their work routines, favoritism, lacunae, and approximations of and deviations from formal policy through the skilled use of rationales, labels, rhetoric, and other language games. The complete web of these learned behaviors, ideas, and words is an *organizational culture*. *Culture* is a tricky term, sometimes insightful but sometimes glib, neutering inquiry. Not all organizations have strong cultures (though some do), not every organization member shares an identical culture or shares it with equal intensity, and organizational cultures are not free-floating isolates but rather are part of wider webs of inequality and power. The best reason to introduce the term *organizational culture* is that, to understand the refraction of official policy into practice, we need to take seriously the everyday working frameworks of bureaucrats. To change policy into practice, then, external mandates or sporadic trainings are not enough—workable new frameworks and reasons to use them need to be introduced. That is a very long and hard process.

CONCLUSION

Anthropological approaches to bureaucracy are diverse, but at their heart, they include certain commonalities. First, they are holistic, seeking to understand bureaucratic work from a variety of perspectives, including those of bureaucrats differently positioned within a workplace hierarchy and across lines of race, ethnicity, class, and gender, and of the publics who interact with the organization.

Second, they are immersive, leveraging the method of participant observation to understand the look and feel of bureaucratic life, including how outcomes of bureaucratic rules are configured by the process of their execution, as well as how informal practices correspond to formal rules.

Third, they envision bureaucracies as social worlds within which a bureaucratic practice is not merely a reflection of bureaucrats' self-interest or psychology but also their socialization into an organizational culture. Whether examining the factors that inform policy implementation, the role of documents in organizing bureaucratic work, or the social role of meetings in the workplace, anthropologists aim to develop a rich account of the multitude of factors that shape how bureaucratic work is understood and carried out.

NOTES

The authors are grateful to Daniel Rogger and other World Bank staff for planning and editorial guidance. The documents section was written by Kristin Asdal and Hilde Reinertsen, and we are grateful to SAGE Publications for allowing us to reprint text from Asdal and Reinertsen's book, *Doing Document Analysis: A Practice-Oriented Method* (2022). The policy section was written by Josiah Heyman. The sections on everyday life and meetings were written by Colin Hoag.

1. For overviews of the anthropological literature on bureaucracies, see Bierschen and Olivier de Sardan (2014, 2021); Heyman (2004); Hoag and Hull (2017). The following are some key anthropological studies of bureaucracy: Ferguson (1994); Gupta (2012); Hetherington (2011); Heyman (1995); Hull (2012a); Mathur (2016); Lea (2008).
2. This section was written by Kristin Asdal and Hilde Reinertsen and builds on chapters 1 and 7 of Asdal and Reinertsen (2022); see also Asdal (2015).
3. Other instructive anthropological studies of the role of documents in bureaucracies are Hetherington (2011); Hull (2012a); Mathur (2016); Riles (2000). For a review of the anthropological literature on bureaucratic documents, see Hull (2012b).
4. These sections are based on chapter 7 of Asdal and Reinertsen (2022).
5. This section was written by Josiah Heyman.

REFERENCES

- Arce, Alberto, and Norman Long. 1993. "Bridging Two Worlds: An Ethnography of Bureaucrat-Peasant Relations in Western Mexico." In *An Anthropological Critique of Development*, edited by Mark Hobart, 179–208. London: Routledge.
- Asdal, Kristin. 2015. "What Is the Issue? The Transformative Capacity of Documents." *Distinktion: Journal of Social Theory* 16 (1): 74–90.
- Asdal, Kristin, and Hilde Reinertsen. 2022. *Doing Document Analysis: A Practice-Oriented Method*. Thousand Oaks, CA: SAGE Publications.
- Bierschenk, Thomas, and Jean-Pierre Olivier de Sardan. 2014. "Studying the Dynamics of African Bureaucracies: An Introduction to States at Work." In *States at Work: Dynamics of African Bureaucracies*, edited by Thomas Bierschenk and Jean-Pierre Olivier de Sardan, 3–33. Boston: Brill.
- Bierschenk, Thomas, and Jean-Pierre Olivier de Sardan. 2021. "The Anthropology of Bureaucracy and Public Administration." In *The Oxford Encyclopedia of Public Administration*, edited by B. Guy Peters and Ian Thynne. Oxford, UK: Oxford University Press. *Oxford Research Encyclopedia of Politics*. <https://doi.org/10.1093/acrefore/9780190228637.013.2005>.
- Brenneis, Donald. 1999. "New Lexicon, Old Language: Negotiating the 'Global' at the National Science Foundation." In *Critical Anthropology Now: Unexpected Contexts, Shifting Constituencies, Changing Agendas*, edited by George E. Marcus, 123–46. Santa Fe, NM: School of American Research Press.
- Brown, Hannah, and Maia Green. 2017. "Demonstrating Development: Meetings as Management in Kenya's Health Sector." *Journal of the Royal Anthropological Institute* 23 (1): 45–62.
- Brown, Hannah, Adam Reed, and Thomas Yarrow. 2017. "Introduction: Towards an Ethnography of Meeting." *Journal of the Royal Anthropological Institute* 23 (1): 10–26.
- Ferguson, James. 1994. *The Anti-Politics Machine: Development, Depoliticization, and Bureaucratic Power in Lesotho*. Minneapolis: University of Minnesota Press.
- Gupta, Akhil. 2012. *Red Tape: Bureaucracy, Structural Violence, and Poverty in India*. Durham, NC: Duke University Press.
- Hetherington, Kregg. 2011. *Guerrilla Auditors: The Politics of Transparency in Neoliberal Paraguay*. Durham, NC: Duke University Press.
- Heyman, Josiah McC. 1995. "Putting Power in the Anthropology of Bureaucracy: The Immigration and Naturalization Service at the Mexico-United States Border." *Current Anthropology* 36 (2): 261–87.
- Heyman, Josiah McC. 2004. "The Anthropology of Power-Wielding Bureaucracies." *Human Organization* 63 (4): 487–500. Republished in *The Anthropology of Organizations*, edited by Alberto Corsin Jimenez, 239–54. Aldershot, UK: Ashgate, 2007.
- Hoag, Colin. 2014. "Dereliction at the South African Department of Home Affairs: Time for the Anthropology of Bureaucracy." *Critique of Anthropology* 34 (4): 410–28.
- Hoag, Colin. 2022. *The Fluvial Imagination: On Lesotho's Water-Export Economy*. Oakland, CA: University of California Press.
- Hoag, Colin, and Matthew Hull. 2017. "A Review of the Anthropological Literature on the Civil Service." Policy Research Working Paper 8081, World Bank, Washington, DC. <https://openknowledge.worldbank.org/handle/10986/26953>.
- Hull, Matthew S. 2012a. *Government of Paper: The Materiality of Bureaucracy in Urban Pakistan*. Berkeley, CA: University of California Press.
- Hull, Matthew S. 2012b. "Documents and Bureaucracy." *Annual Review of Anthropology* 41 (1): 251–67.
- Lea, Tess. 2008. *Bureaucrats and Bleeding Hearts: Indigenous Health in Northern Australia*. Sydney: University of New South Wales Press.
- Mathews, Andrew S. 2011. *Instituting Nature: Authority, Expertise, and Power in Mexican Forests*. Cambridge, MA: MIT Press.
- Mathur, Nayanika. 2016. *Paper Tiger: Law, Bureaucracy and the Developmental State in Himalayan India*. Cambridge, UK: Cambridge University Press.
- Olivier de Sardan, Jean-Pierre. 2009. "State Bureaucracy and Governance in Francophone West Africa: An Empirical Diagnosis and Historical Perspective." In *The Governance of Daily Life in Africa: Ethnographic Explorations of Public and Collective Services*, edited by Giorgio Blundo and Pierre-Yves Le Meur, 39–72. London: Brill.
- Riles, Annelise. 2000. *The Network Inside Out*. Ann Arbor, MI: University of Michigan Press.
- Schwartzman, Helen B. 1989. *The Meeting: Gatherings in Organizations and Communities*. New York: Springer.
- Thedvall, Renita. 2013. "Punctuated Entries: Doing Fieldwork in Policy Meetings in the EU." In *Organisational Anthropology: Doing Ethnography in and among Complex Organisations*, edited by Christina Garsten and Anette Nyqvist, 106–19. London: Pluto Press.