



GOVERNANCE

GOVERNANCE

**EQUITABLE GROWTH, FINANCE & INSTITUTIONS INSIGHT**

## **Public-Sector Productivity (Part 1)**

# Why is it important and how can we measure it?

February 2021

Ravi Somani  
Bureaucracy Lab  
Development Impact Evaluation Group  
and Governance Global Practice



WORLD BANK GROUP

Public Disclosure Authorized

Public Disclosure Authorized

Public Disclosure Authorized

Public Disclosure Authorized

Some rights reserved.

This work is a product of the staff of The World Bank with external contributions. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Nothing herein shall constitute or be considered to be a limitation upon or waiver of the privileges and immunities of The World Bank, all of which are specifically reserved.

#### Rights and Permissions



This work is available under the Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO), <http://creativecommons.org/licenses/by/3.0/igo>. Under the Creative Commons Attribution license, you are free to copy, distribute, transmit, and adapt this work, including for commercial purposes, under the following conditions:

**Attribution**—Please cite the work as follows: 2021. “Public-Sector Productivity (Part 1) Why is it important and how can we measure it?” EFi Insight-Governance. Washington, DC: World Bank.

**Translations**—If you create a translation of this work, please add the following disclaimer along with the attribution: *This translation was not created by The World Bank and should not be considered an official World Bank translation. The World Bank shall not be liable for any content or error in this translation.*

**Adaptations**—If you create an adaptation of this work, please add the following disclaimer along with the attribution: *This is an adaptation of an original work by The World Bank. Views and opinions expressed in the adaptation are the sole responsibility of the author or authors of the adaptation and are not endorsed by The World Bank.*

**Third-party content**—The World Bank does not necessarily own each component of the content contained within the work. The World Bank therefore does not warrant that the use of any third-party-owned individual component or part contained in the work will not infringe on the rights of those third parties. The risk of claims resulting from such infringement rests solely with you. If you wish to reuse a component of the work, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright owner. Examples of components can include, but are not limited to, tables, figures, or images.

All queries on rights and licenses should be addressed to World Bank Publications, The World Bank Group, 1818 H Street NW, Washington, DC 20433, USA; e-mail: [pubrights@worldbank.org](mailto:pubrights@worldbank.org).

Graphic Designer: Diego Catto / [www.diegocatto.com](http://www.diegocatto.com)

This note is the first of a two-part series that explores the importance of public-sector productivity and its measurement (part 1); and its determinants (part 2).

This note summarizes a review of the literature on different approaches to measuring public-sector productivity (the rate at which inputs are converted into outputs).

Accurately measuring public-sector output and productivity is important for four primary reasons:

- i. Public employment accounts for 38% of formal employment globally, making it an important contributor to national measures of labor productivity.
- ii. Public expenditure accounts for a quarter of GDP globally and the wage bill around 8% of GDP. As such, improvements in public-sector productivity can have substantial implications for public finances, including development-assistance funds.
- iii. Core indicators of social welfare, such as the [Sustainable Development Goals](#) and the [World Bank's Human Capital Index](#), rely on a well-functioning public sector. Accurately measuring the efficiency and quality of public-service provision is, therefore, necessary to track progress against development targets.
- iv. Accurate measures of public-sector productivity allow senior officials to monitor, benchmark, and incentivize agencies and employees and better allocate scarce resources.

However, accurately measuring inputs and outputs in the public sector can be challenging for several reasons, including: the lack of market transactions for public goods and services; the existence of market imperfections in the sectors in which the public sector typically operates; and the complex nature of public services, requiring multiple, hard-to-measure inputs from multiple sources, and exhibiting a time lag between investment in inputs and the realization of outputs.

Drawing on the literature, this note presents different approaches to measuring public-sector productivity and a discussion of the advantages and disadvantages of each, summarized in Table 1. Table 2 summarizes the technical and data requirements for implementation of each approach.

This note recommends:

1. Complementing traditional 'macro' measures of public-sector productivity, such as the cost-weighted-output approach presented in [Atkinson \(2005\)](#), with fine-grained 'micro' measures at the individual organization, employee, and task/process level.
2. Monitoring and reporting output (performance) measures and inputs (costs) separately.
3. Combining multiple measures of productivity, tied closely to the service-delivery chain.



# Why Should We Measure Public-Sector Productivity?

Productivity is a central concept in the economics literature that extends from the idea that the output produced by an organization or an economy is some function of the available inputs, such as labor, capital, and technology ([Solow, 1957](#)). Specifically, productivity measures the efficiency with which a given organization transforms inputs into outputs. Productivity growth is seen as a central driver of firm profitability and economic growth ([Romer, 1990](#); [Mankiw et al., 1992](#); [Foster et al., 2008](#); [Bloom et al., 2013](#)). In fact, Paul Krugman, winner of the 2008 Nobel Prize for Economics, famously stated that “*Productivity isn’t everything, but, in the long run, it is almost everything. A country’s ability to improve its standard of living over time depends almost entirely on its ability to raise its output per worker*” in his book titled ‘[The Age of Diminished Expectations](#)’.

Generating productivity growth allows organizations to produce more desired outputs for the same quantity of inputs.<sup>1</sup> In the private sector, this means greater profitability and firm growth. In the public sector, this means improved levels of service delivery and indicators of social welfare for the same level of expenditure or, more generally, more efficient and impactful public expenditure. **Generating productivity growth – doing more with less – is likely to remain a priority for governments in the post-pandemic period.**

Accurate measures of productivity allow firms, organizations, and policymakers to keep track of how individual inputs, such as labor, are being converted into desired outputs. With such information, organizations and policymakers can identify and attempt to address any downturns in productivity and better identify which technologies, policies, and practices are productivity-enhancing.

**Below, we highlight four main reasons why accurately measuring output and productivity in the public sector is important.**

---

1. Or the same amount of desired output with fewer inputs.

# The Measurement of National Labor Productivity

The public sector is a major employer in most economies. Globally, the public sector accounts for 16% of total employment and 38% of formal employment according to the [Worldwide Bureaucracy Indicators \(WWBI\)](#), published by the World Bank's Bureaucracy Lab (see figure 1). In OECD countries, public employment accounts for 20% of total employment. Measuring the productivity of such a sizeable proportion of the workforce is important for understanding the trends in and determinants of national labor productivity.

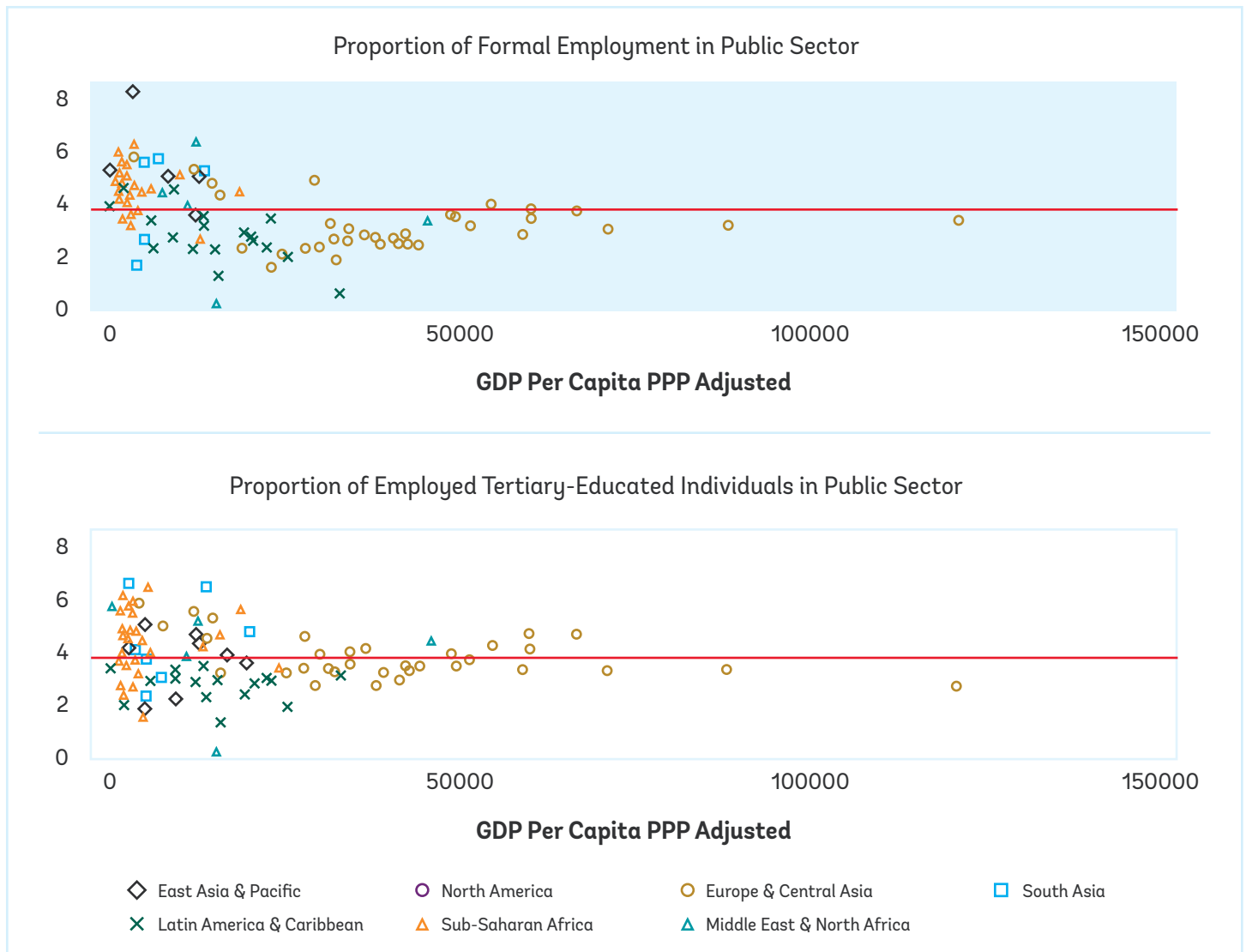
Further, the public sector is an especially important employer for high-skilled labor: 39% of all tertiary-educated employment

is within the public sector according to the WWBI (see figure 1). Hence, accurate measures of public-sector productivity are necessary to shed light on the labor-market implications of education programs.

Accurate measures of labor productivity allow governments to compare the efficiency of labor and productivity trends across different sectors of the economy. This information can improve our understanding of the drivers of labor productivity and the types of policies that can generate productivity growth in the economy.

> > >

**FIGURE 1 - Public employment as a proportion of formal employment (top). Public employment as a proportion of all tertiary-educated employment (bottom). The red line represents the global average.**



Source: *Worldwide Bureaucracy Indicators*





## The Management of the Public Purse (Including Development Assistance)

---

The public sector is a major purchaser of goods and services in the economy. Government expenditure accounts for 27% of GDP globally and 42% of GDP in OECD countries. The wage bill of the public sector accounts for 8% of GDP globally according to the WWBI. Small improvements in the efficiency with which the public sector converts expenditures into outputs holds the potential to widen fiscal space and free up public resources for priority investment areas. Similarly, development-assistance and aid projects depend on government agencies for coordination and [implementation](#). As such, the efficiency and productivity of public-sector organizations in recipient-country governments is an important determinant of the efficiency and success of development-assistance and World Bank projects.

Productivity data for the public sector allows policymakers to identify high- and low-productivity individuals and organizations and pinpoint opportunities where productivity improvements are possible. Accurate measures of public-sector output and productivity also allow policymakers and citizens to better understand the costs and benefits of investments made in the public sector. This can lead to improved knowledge over which types of investments and reforms yield the greatest productivity gains and why they were (not) successful. For example, governments have made large investments in IT infrastructure and GovTech, but there is little evidence on the impacts of such investments on output or productivity in the public sector ([World Development Report, 2016](#)). These opportunities are particularly relevant for the functioning of resource-strained governments in the pandemic and post-pandemic eras and their ability to effectively meet citizen needs.

## The Monitoring, Benchmarking, and Incentivization of Public Agencies and Officials

---

As noted above, the public-sector wage bill accounts for 8% of GDP globally and the public-sector workforce accounts for 38% of formal employment. Ensuring that the taxpayer gets the most 'bang for their buck' means giving public-sector managers the necessary tools to monitor and incentivize the performance of public employees. **Individual- and organization-level productivity data can help managers and employees identify new innovations, reforms, and best practices that improve productivity.** Such data can provide useful internal benchmarking and monitoring tools, as well as effective incentive devices when linked to rewards systems. Furthermore, such data may provide useful insights into cross-national comparisons of common public-sector organizations. As such, accurately measuring public-sector productivity can generate productivity growth in and of itself.

## The Achievement of Development Goals and Social Welfare Objectives

---

Many core indicators of social welfare, such as those included in the [Sustainable Development Goals](#) and the [World Bank's Human Capital Index](#), covering education, health, clean water, clean energy, infrastructure, reduced inequalities, climate action and conservation, and justice, are highly dependent on a well-functioning public sector. Accurately measuring the quality and efficiency of public-sector outputs is vital for tracking progress towards important development objectives and identifying barriers towards their achievement.



# The Challenges of Measuring Public-Sector Productivity

Productivity is a measure of the efficiency with which inputs (labor, capital and raw materials) are converted into outputs. The standard definition of total-factor productivity is: the ratio of the currency-value of outputs to the currency-value of all inputs. Often, studies of labor markets and firms focus specifically on labor productivity: the ratio of the currency-value of outputs to the currency-value expended on all labor inputs.

$$\text{Total Factor Productivity} = \frac{\text{Outputs (USD)}}{\text{Inputs (USD)}}$$

$$\text{Labor Productivity} = \frac{\text{Outputs (USD)}}{\text{Expenditure on Labor (USD)}}$$

Some examples of productivity measures include:

- The value (revenue) of shoes produced per dollar spent on total labor inputs (labor productivity, as output per unit of expenditure on labor)
- The value (revenue) of shoes produced per dollar spent on labor, machines, and raw materials (total factor productivity)

**Accurate measures of productivity require reliable measures of the value of inputs and outputs.** However, obtaining accurate valuations of inputs and outputs can be challenging for many sectors, services, and goods in the economy ([Syverson, 2011](#)). For example, measuring the output value of services can be difficult because service transactions are often heterogenous on certain margins, such as the nature of the interaction between provider and consumer or the specificities of the consumers' case or request ([Griliches, 1992](#)). This heterogeneity creates difficulties in separating variations in quantity from quality and determining unit prices of comparable service outputs. Many services also exhibit asymmetric information between the provider and consumer, such as those provided by lawyers, doctors, and teachers, which means that consumers are not able to accurately estimate the value of the services that they receive. Such asymmetric information leads to inaccurate estimates of output value based on market prices or consumer estimates.

The growth of digital services has added further challenges to measuring service-sector productivity. For example, many digital services are characterized by complex production functions and supply chains; are provided to consumers for free as part of a 'two-sided market'; and exhibit substantial product-, firm- and time-variation in quality ([Rysman, 2009](#); [Coyle, 2018](#); [Brynjolfsson et al, 2018](#); [Coyle, 2019](#); [Coyle and Nguyen, 2019](#)). The first of these conditions makes it difficult to identify, value and attribute individual inputs to particular products, outputs, or activities. The second and third make it difficult to accurately determine the value of the output and adjust valuations over time to account for differences in service quality. Such factors, and more, are apparent in the public sector. **We summarize the seven major challenges of measuring public-sector productivity below.**

## Seven Challenges of Measuring Public-Sector Productivity

### Measuring Outputs

**First, there are no market transactions for most government services; and where they do exist, they are heavily distorted by subsidies.** The lack of (competitive) market transactions makes it difficult to determine the economic value of a public service, as consumers cannot reveal their valuations through quantities purchased or prices paid. This makes the value of outputs difficult to elicit.

**Second, many public goods and services are provided because of market imperfections.** For example, many public goods and services exhibit positive externalities or asymmetric information between the provider and consumer. The very nature of such goods and services means that market-based valuations of output may be inaccurate. This is because consumers (or regulators) do not fully account for the social benefit of such goods and services (in the case of externalities) or because consumers (or regulators) have insufficient information to accurately value the quantity and quality of such goods and services (in the case of asymmetric information). For example, in the health sector, citizens may not account for the social value of a vaccine, limiting demand; or citizens may be unable to distinguish between high- and low-quality healthcare services due to a limited understanding of complex health procedures. Using market-based measures of outputs, such as patient numbers, satisfaction ratings, provider revenue, or even measurable health indicators may lead to inaccurate assessments of provider productivity.

**Third, in several sectors where the government is the main provider, the services are collective in their nature and outcomes are broad or difficult to measure.** For example, in defence, foreign relations, and environmental protection, outcomes can be difficult to identify and measure even at the national level; and services are provided to and consumed collectively by the nation as a whole, rather than any specific or identifiable household or locality. This makes the output of public-sector organizations operating in these sectors difficult to identify, measure, and benchmark.<sup>2</sup>

**Fourth, public provision is tied to the wider political economy of the public sector.** The political underpinnings of public-service delivery can generate additional frictions in the markets for public services that increase the difficulty of measuring public-sector output. For example, politically determined boundaries can limit consumers' access to public services based on geography or socio-economic characteristics, limiting the efficacy of demand-based measures of public-sector outputs (e.g. usage rates and provider revenue). Politically determined barriers can also limit the entry of potential providers into the market, limiting the utility of market-based measures of public-sector output (e.g. provider revenue and growth). Similarly, budget processes that are politically determined and/or unrelated to efficiency and performance considerations limit the potential for market-based measures of output value. In addition, public-sector wages that are politically determined and de-coupled from workers' skills and measures of performance impede the use of market-based approaches to measuring labor-productivity (e.g. wage residuals).

**Fifth, the quality of public services changes over time.** Not adjusting for changes to quality can lead to inaccurate estimates of the value of output. For example, certain healthcare procedures have improved significantly over time in terms of their safety and effectiveness. Measures of output that do not account for this such improvements in quality will understate the value of the output produced.

### Measuring Inputs and Correctly Attributing Inputs to Outputs

**Sixth, many public goods and services are relatively complex, requiring multiple (often immeasurable) inputs from multiple individuals and sources** ([Dixit, 2002](#)). For example, [this](#) OECD review documents how public-sector functions increasingly involve a range of internal and external agents (contractors) and how client-facing public employees have to provide direct services to clients as well as support clients to navigate the "complex webs of services, entitlements, bene-

2. For example, even if one is able to define a different set of outputs for the organization, such as those based on processes or tasks, there is no way of validating whether the outputs are associated with positive or negative outcomes.





fits and eligibility requirements”. The UK government’s [Annual Report on Major Projects](#) details public infrastructure projects that are particularly complex, high-risk, and/or innovative, requiring a range of different areas of expertise and involving numerous stakeholders and providers. Even the preparation of a seemingly simple monitoring or budget report can require the involvement of several public officials, the exact contributions of which are difficult to measure and cost.

This multi-dimensionality and complexity of the production function for public services makes it difficult to accurately measure and attribute the individual inputs used in the production of specific outputs.

**Seventh, there is often a time lag between investments in public-service inputs and the production of outputs or outcomes.** For example, investments in new healthcare procedures, including the research, testing and approval processes, can occur many years or decades before the services are widely available. Moreover, any effects on health outcomes may take many more years to transpire. As such, attributing individual expenditures to specific outputs or outcomes becomes difficult in practice. Other areas where there is often a significant time lag between investments in inputs and the production of outputs or outcomes include education and durable public infrastructure.

## Performance or Productivity?

The major difference between measuring productivity and measuring performance is that the former attempts to take into account expenditure on inputs. If conducted accurately, this is preferable because achieving substantial gains in performance may be undesirable if the gains occur only as a result of excessive increases in input expenditure. However, in cases where the production function is complex and where quality is difficult to capture, a focus on singular productivity metrics can obscure important underlying trends in different components of an organization’s operations. In addition, the use of productivity metrics in benchmarking and monitoring exercises may incentivize excessive cost reduction at the expense of harder-to-capture service quality, especially when there is a time lag between input expenditures and their measurable effects on services.

For example, a rapid increase in productivity metrics such as the number of healthcare procedures performed per dollar of input expenditure may arise due to: (i) an increase in the number of procedures performed per worker, holding costs and quality constant; (ii) a reduction in assurance and monitoring standards (a decline in quality); (iii) a decrease in input expenditures, such as labor costs, equipment-rental costs, or contracting costs. Clearly, (i) is desirable, (ii) is potentially undesirable, and (iii) is ambiguous, depending on whether the decrease in input expenditure leads to reductions in the quality of inputs (e.g. the qualifications and motivation of staff, the quality of contracted services, the quality of equipment, etc) that manifest into future declines in output quality.

Capturing both productivity and performance – separately tracking and reporting each component of the productivity-metric calculation – will allow for a better understanding of the drivers of public-sector productivity and minimize potentially harmful adverse incentives.



# A Summary of Measurement Approaches

## The Conventional Approach: Inputs as Outputs

---

Due to the difficulty of measuring public-sector outputs, described above, the conventional approach to measuring public-sector productivity has been to use expenditure on inputs instead of outputs. For example, until 1998 the UK government used measures of employee compensation, the procurement cost of goods and services, and a charge for the consumption of fixed capital in the national accounting as measures of government output. However, since these are inputs, and the standard definition of productivity is the ratio of outputs to inputs, this approach assumes that the productivity of the public sector is always equal to one – it never grows and never declines!

## Direct Measures of Public-Sector Productivity

---

More recently, alternative approaches to directly measuring government output have been used, particularly within 'individual' service sectors, such as education, health, tax, customs, and social security.

This section presents a brief summary of the direct measures of public-sector productivity. For simplicity, each measure is broadly categorized into 'macro' and 'micro' and summarized in Table 1.

Macro approaches are defined as those that can only provide aggregate information – at the level of an organization, sector, or the whole public service. These are the types of approaches traditionally used to measure public-sector productivity, including the cost-weighted productivity measures proposed in [Atkinson \(2005\)](#) and [Dunleavy \(2017\)](#), as well as service-delivery indicators and budget-execution rates. Such measures cannot be attributed to individual employees.

Micro approaches are defined as those that can extend down to individual employees, tasks, and processes. These include measures of process-completion rates, task-completion rates, procurement outcomes (such as unit prices for the same good), measures of staff and user satisfaction, subjective performance assessments and evaluations by key stakeholders and employees, measures of process productivity or the use of independent observers of standardized public-service requests, and measures of the knowledge, skills and capacity of public officials.

**TABLE 1 - Summary of measures of public-sector output (advantages and disadvantages)**

Macro (organization, sector, whole public service)	Micro (employee, task, process)
<p style="text-align: center;"><b>Cost-weighted output (Atkinson, 2005)</b></p> <ul style="list-style-type: none"> <li>✓ closely tied to the definition of productivity</li> <li>✓ useful to track productivity trends</li> <li>✗ absolute levels are difficult to interpret</li> <li>✗ the approach will vary across sectors</li> <li>✗ needs activity-level data on unit costs</li> <li>✗ difficult to incorporate changes in quality</li> </ul>	<p style="text-align: center;"><b>Process-, project- and task-completion rates</b></p> <ul style="list-style-type: none"> <li>✓ applicable to a broad range of services and settings</li> <li>✓ allows the identification of good and bad performers down to individual level and correlates of performance</li> <li>✗ difficult to capture process quality and complexity</li> <li>✗ completion rates depend on quality of initial plan</li> <li>✗ may incentivize low-quality, low-complexity tasks</li> </ul>
<p style="text-align: center;"><b>Service delivery indicators</b></p> <ul style="list-style-type: none"> <li>✓ closely linked to welfare outcomes</li> <li>✓ provides sector-specific productivity measures</li> <li>✗ more appropriate for frontline services</li> <li>✗ may depend on external (e.g. demand-side) factors</li> <li>✗ may incentivize attention to certain indicators</li> </ul>	<p style="text-align: center;"><b>Procurement outcomes</b></p> <ul style="list-style-type: none"> <li>✓ allows the identification of good and bad performers down to individual level and correlates of performance</li> <li>✗ specific subset of public-sector activities</li> <li>✗ difficult to capture variations in quality</li> <li>✗ external factors determine local prices</li> </ul>
<p style="text-align: center;"><b>Budget execution rates</b></p> <ul style="list-style-type: none"> <li>✓ simple and readily available data across most organizations/settings</li> <li>✗ may not indicate a well-functioning organization</li> <li>✗ relies on the quality of budget-preparation process</li> </ul>	<p style="text-align: center;"><b>Staff and user satisfaction</b></p> <ul style="list-style-type: none"> <li>✓ allows the identification of good and bad performers down to individual level and correlates of performance</li> <li>✗ relies on prior expectations of users and staff</li> <li>✗ difficult to capture variations in case complexity</li> <li>✗ user satisfaction only feasible in a subset of offices</li> <li>✗ staff satisfaction may capture pay, conditions, etc</li> <li>✗ may be biased by political preferences</li> </ul>
	<p style="text-align: center;"><b>Subjective assessments (by employees, by stakeholders)</b></p> <ul style="list-style-type: none"> <li>✓ allows the identification of good and bad performers down to individual level and correlates of performance</li> <li>✓ applicable to a broad range of services and settings</li> <li>✗ stakeholder characteristics/preferences may vary</li> <li>✗ use of appraisals depends on quality of process</li> <li>✗ staff assessments may capture pay, conditions, etc</li> </ul>
	<p style="text-align: center;"><b>Independent observers ('mystery shoppers') and process productivity</b></p> <ul style="list-style-type: none"> <li>✓ allows the identification of good and bad performers down to individual level and correlates of performance</li> <li>✓ applicable to a broad range of services and settings</li> <li>✗ can be costly to implement</li> <li>✗ needs set of comparable best practices (checklists)</li> </ul>
	<p style="text-align: center;"><b>The knowledge/capacity of public officials</b></p> <ul style="list-style-type: none"> <li>✓ allows the individual-level measures</li> <li>✓ applicable to a broad range of services and settings</li> <li>✗ information does not necessarily translate to output</li> <li>✗ certain areas of knowledge are difficult to measure</li> </ul>

# Moving Towards Micro-Level Approaches

**A primary motivation for using micro-level measures of public-sector productivity is based on the observation that performance varies substantially within and across public-sector organizations.** For example, in many of the settings in which the Bureaucracy Lab has collected data on management practices or micro-level measures of productivity, the types of patterns presented in figures 2 and 3 are commonly found. This mirrors a similar trend observed in the private sector, where productivity varies significantly between firms and plants within countries and even within narrowly defined sectors ([Bloom and Van Reenen, 2007, 2010](#)).

This distribution of performance means that one can simply walk across the corridor in a given organization (where

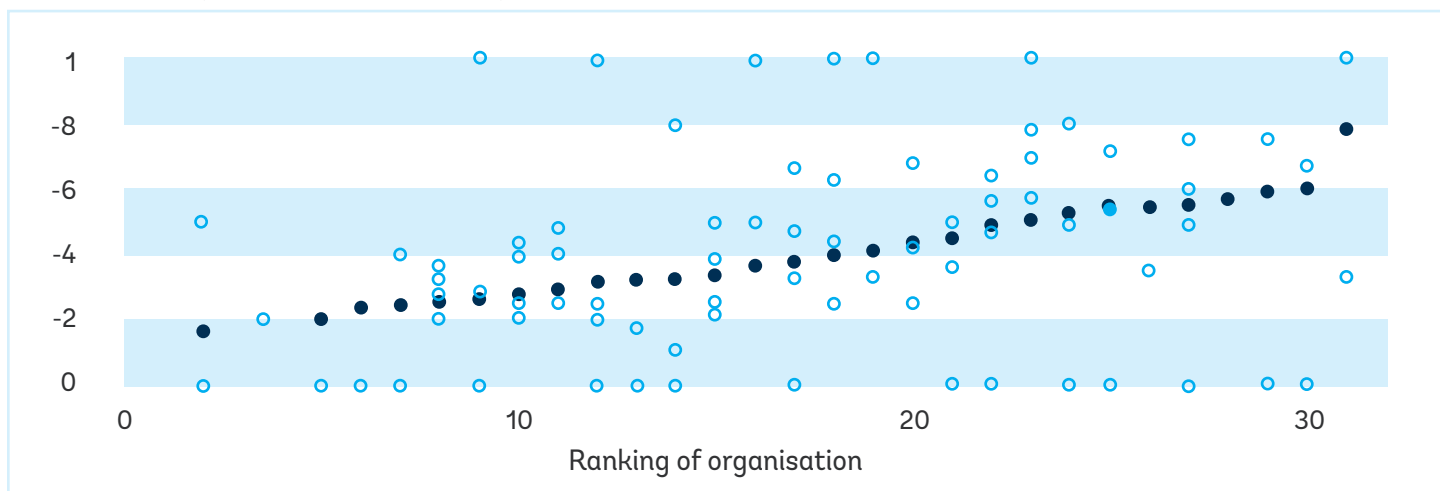
the characteristics of the geography, political environment, senior management, citizens, etc, are held constant) and find substantially different rates of performance and management quality.

**Such patterns imply that it is difficult to characterize the productivity of even individual public-sector organizations, let alone whole sectors, regions, or administrations.**

This fact lessens the utility of broader (macro) measures of productivity for tracking public-sector performance and for managerial purposes, such as identifying best practices and successful reforms, as well as monitoring, benchmarking and incentivizing the performance of public officials.

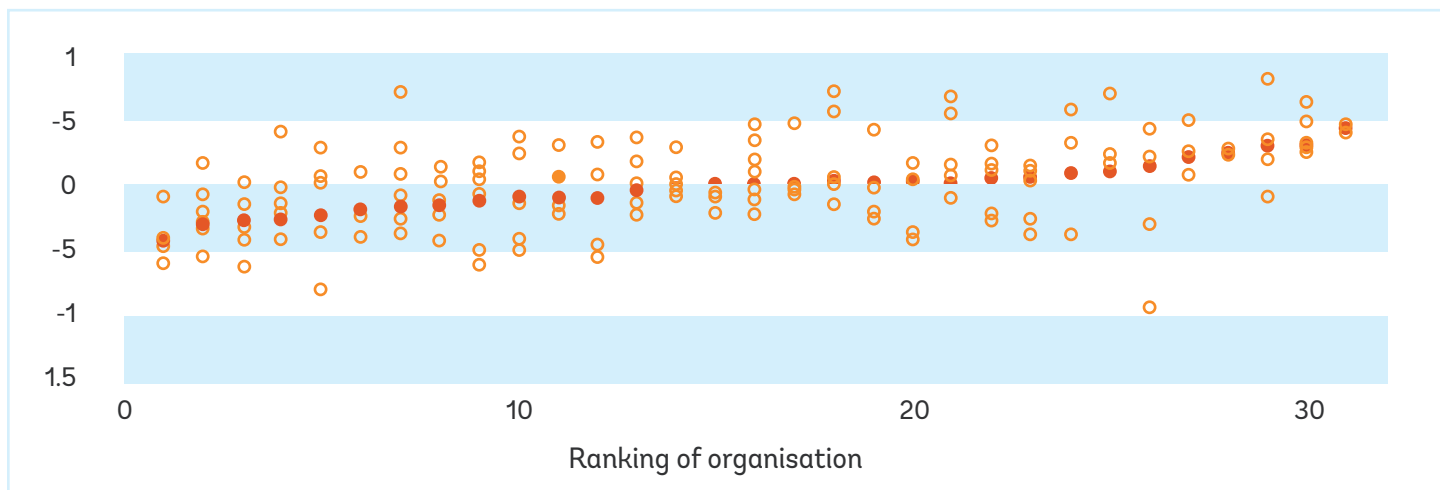
> > >

**FIGURE 2 - Project Completion Rates for Units (Light Blue, Hollow Circles) and Organizations (Dark Blue, Solid Circles) in the Ghanaian Civil Service**



> > >

**FIGURE 3 - Management Practices for Units (Light Orange, Hollow Circles) and Organizations (Dark Orange, Solid Circles) in the Ghanaian Civil Service**



**In addition, capturing micro-level measures of performance and productivity increases the variation and data available to identify the correlates and potential determinants of public-sector productivity.** This information, when combined with further micro-level data from surveys and administrative sources, allows researchers to begin studying the following types of questions: which characteristics of workers are associated with better performance; which management practices incentivize better performance; how important are certain resources for public-sector productivity? A similar path of discovery in relation to private-sector productivity, driven by micro-empirical evidence, led to a range of important findings about its potential determinants ([Bloom and Van Reenen, 2007, 2010](#); [Bandiera et al., 2011](#); [List and Rasul, 2011](#); [Bloom et al. 2013](#); [Konings and Vanormelingen, 2015](#); [Lazear, 2018](#); [Ashraf and Bandiera, 2018](#)). This micro-empirical approach is growing in importance in research on the public sector ([Finan et al., 2017](#)).

Micro-level data and the variation that comes with it also creates opportunities for impact evaluations within public-sector organizations to causally measure the effectiveness of reforms on public-sector productivity. For example, do financial incentives generate performance gains; do performance-based transfers work; do bonuses improve performance; what is the potential effectiveness of e-procurement on unit prices; can recruitment technologies attract motivated and skilled employees and improve performance? There are many examples of such work in the private sector and increasingly in the public sector ([Bandiera et al., 2011](#); [List and Rasul, 2011](#); [Finan et al., 2017](#)).





# Direct Measures of Public-Sector Output: Details

## Macro: Cost-weighted Output

This approach builds on [Atkinson \(2005\)](#) and applies the following steps to develop organization- and service-level measures of productivity ([Dunleavy, 2017](#)):

1. Identify core organizational activities and define unit costs for each.<sup>3</sup>
2. Develop a cost-weighted total output measure for the agency, which can be done by the following type of calculation:
  - Cost-weighted total output = (units of activity1 \* unit costs for activity1)+(units of activity2 \* unit costs for activity2) +...(units of activityN \* unit costs for activityN)
3. Calculate total input expenditure, accounting for all of the inputs used to make all of the activities included in the output calculation
  - Input expenditures include: total labor costs, including externally sourced and temporary labor; intermediate costs, such as intermediate outputs from outsourced activities; property costs; procurement expenses; maintenance costs; etc.

These measures are then converted into a productivity metric for the organization or service by dividing cost-weighted output (2) by total input expenditure (3). To account for inflation in unit costs and input expenditures over time, the calculation should deflate unit costs and input expenditures by an appropriate measure of inflation: this should be appropriate to the sector, product and/or service in question. To account for changes in service quality over time, quality measures or proxies can be used to adjust the output measure (e.g. using data on complaints; response times; official star ratings; infrastructure quality measures; surveys of user experience; staff perceptions).

The advantage of this approach is that it is closely related to the definition of productivity and can be used for tracking and benchmarking changes in productivity over time for the public-sector as a whole, as well as for individual sectors.

However, the exact approach may vary across sectors and organizations due to differences in activities, the types of inputs, or data availability, making it difficult to interpret differences in productivity measures across sectors and organizations. In addition, while the measure is useful for tracking changes in productivity over time, the absolute values are often difficult to interpret and utilise for policy decisions.

3. In private-sector organizations, prices are used in the calculation of revenue. In the public sector, with the absence of market prices, unit costs are used instead.

To implement this approach in practice, one needs to identify core activities for organizations, which can be difficult, especially for collective public services. This approach also requires detailed data on the individual inputs used for each activity and their costs. For example, how much of total labor expenditure went into the specific activity used in the calculation? In addition, some inputs and input costs are difficult to identify and take into account (e.g. IT systems; durable goods over years; wear & tear; inherited land; etc).

To allow comparability over time, inflation and changes in output quality need to be accounted for. There are different measures of inflation to choose from, which can be economy-wide, sector-specific, or based on a select basket of goods and services. Choosing the most appropriate measure of inflation is important to analyse temporal productivity trends. Quality is typically difficult to measure and capture in a holistic manner. When only certain dimensions of service quality are measurable and tracked, providers may be incentivized to focus on this set and neglect other (less salient) dimensions of quality ([Dixit, 2002](#)).

## Macro: Service Delivery Indicators

---

This approach involves using service delivery indicators, such as core health and education indicators, as proxies for public-service output. These are typically only available and appropriate for certain frontline service providers rather than those in middle-tier or centralized levels of the public sector. Examples include the use of indicators of health outcomes in [Propper et al \(2007\)](#), the use of education indicators in [Muralidharan and Sundararaman \(2011\)](#), and indicators of employment in public-works programs in [Gulzar and Pasquale \(2017\)](#)

These measures align closely with indicators of social welfare and may therefore be of key public interest. However, this measurement approach is typically only relevant for organizations performing frontline work. Even then, many service delivery indicators depend on factors outside of the control of public officials, such as demand-side, environmental, and political factors, which may make it difficult to distinguish between productivity gains in local public organizations and external local changes.

The use of such indicators for measuring output and productivity may also incentivize public resources and aggregate effort towards these measurable aspects of service delivery at the expense of other, less-salient dimensions.

## Macro: Budget Execution Rates

---

This approach uses the proportion of budget that is actually disbursed or expended as a measure of efficiency. An example of this approach is [Liebman and Mahoney \(2017\)](#), which uses the percentage of budget spent over the fiscal year and makes adjustments for quality.

While this approach may be simple and feasible across many contexts and sectors, it has two important shortcomings: (i) it relies on the initial allocation of the budget being fairly and reasonably distributed; (ii) fast spending is not necessarily consistent with quality or efficient service provision. For example, a low budget-execution rate could indicate that an organization is experiencing implementation challenges or that an organization has discovered efficiency savings, producing the expected levels of service with lower-than-planned expenditure levels. The use of such an approach could also incentivize rapid fund-disbursement in the long run, which may come at the expense of careful, diligent, socially optimally allocation decisions.

## Micro: Task Completion Rates and Individual Performance Indicators

---

This approach measures how many projects, processes, cases, or tasks are completed and, in some cases compares this to a pre-specified plan or commitment. Examples include using unit-level project-completion rates for local public-infrastructure projects ([Rasul and Rogger, 2018](#)), unit-specific task-completion rates relative to an annual plan across various tasks in the federal civil service ([Rasul et al. 2020](#)), organization-level road-construction-completion information ([Lewis-Faupel et al. 2016](#)), individual employee-level case-completion times in the social-security administration ([Fenizia, 2020](#)), and organization-level tax-audit completion rates ([Dunleavy and Carrera, 2013](#)). With fine-grained data, such approaches can be taken down to the individual level, providing rich variation for analyses of correlates.

Other examples of individual-level measures of the performance or output of public servants, include household visits by individual healthcare workers ([Ashraf et al. 2014](#)), tax revenues collected by individual tax-agency employees ([Khan et al. 2019](#)), and attendance rates and hours worked by health officials ([Callen et al. 2020](#)).

Such approaches may be susceptible to unobservable variation in the quality and complexity of individual processes and tasks. For instance, without good measures of quality and complexity, it is difficult to say whether a fast completion rate is really a good outcome. Measures that rely on completion rates relative to a pre-specified plan or commitment also depend on the quality of the initial plan. For example, an unambitious plan allows for a high project-completion rate even among poor performers.

If such measurement issues are taken into account, micro-level data on process completion allows public-sector managers to more accurately identify good and bad performers and determine the personnel practices that are associated with good performance.

However, in the long run the use of such measures can incentivize the reduction of targets or a re-allocation of resources towards easier-to-complete and less complex tasks.

## Micro: Procurement Outcomes

---

Procurement is a sub-sector of the public administration that has a ready-made productivity measure: unit prices paid for the same product. With fine-grained data on product specifications – the specific type of product and measures of its quality – combined with price paid and quantity received, procurement systems can generate data on the unit prices that individual procurement officers and organizations pay for the same good. Examples of this approach include [Lewis-Faupel et al \(2016\)](#) and [Best et al \(2019\)](#).

However, such measures rely on being able to accurately categorise procured products in way that fully captures differences in product specifications and measures of quality. Furthermore, there may be external factors that determine the price of procurement paid by a given officer or organization, such as the local product market or transport costs, that need to be taken into account for fair comparisons of performance across individuals and entities.

## Micro: User and Staff Satisfaction Rates

---

This approach takes measures of user satisfaction or staff/employee satisfaction, based on surveys, as a measure of

performance. Examples include villager satisfaction with current levels of public services and their willingness to contribute to future services ([Olken, 2010](#)), or satisfaction rates of users of local public services ([Boyne, 2002](#); [Andrews et al, 2005](#)). Examples also include measures of staff satisfaction or current levels of motivation ([Janke et al, 2019](#)) or objective measures of staff satisfaction, such as turnover or retention rates, particularly of key staff ([Bright, 2008](#); [Deserranno, 2019](#)).

The downsides of user-satisfaction measures concern the issues of case complexity, differential user expectations and preferences, and internal biases. For example, a complex case may be difficult for public servants to address, even if the public servant performs to a high level and is highly capable. Citizens may demonstrate different preferences and priorities over features of the service (e.g. quantity versus quality) or have different prior expectations of service quality or quantity, impacting how they evaluate the performance of the provider. Satisfaction measures may also capture other internal biases held by the citizen, for example their political preferences and general attitudes to the incumbent government. In such instances, some public servants may receive lower average satisfaction scores even when their levels of effort and performance are relatively high.

Another concern is that user satisfaction rates often bundle together various different margins of satisfaction (e.g. the politeness of the public servant, the professionalism and knowledge of the public servant, the characteristics of the case, the office environment, the time of day/year, etc) that make it difficult for managers to identify specific actions to take. Further, user-satisfaction may only be appropriate in a subset of public-sector offices: those with users.

Staff satisfaction rates are also likely to be capturing a whole range of issues unrelated to organizational performance, including workers' satisfaction with pay, working conditions, relationship with management, mission alignment, etc, as well as any internal biases of workers, such as those relating to political preferences.

## Micro: Subjective Assessments (Employees and Stakeholders)

---

This approach involves directly asking employees and/or important stakeholders how they perceive the performance of their own division/unit/organization or their respective local government. One approach includes the use of employee surveys that ask about different aspects of organizational perfor-

mance (Brewer and Selden, 2000). A potentially more holistic approach, used by [Bertrand et al \(2018\)](#), involves surveying a multitude of important stakeholders, such as managers, civil servants, politicians, business associations, local media, and civil society organizations to elicit their subjective assessments of the performance of their local government. A third type of approach within this realm involves using the performance-evaluation data that arises from the annual appraisals of public-sector employees occurring in most administrations ([Alonso and Lewis, 2001](#); [Somani, 2020](#)). This provides information on the managers' subjective assessments of their employees.

While such measures provide individual-level measures of performance, they encompass the types of measurement bias that arises with perception surveys and subjective assessments. For example, a survey of employees' perception of their own organization's performance is likely to also capture their dissatisfaction with their working conditions, such as pay, their manager, or other potentially non-performance-related variables. Using subjective assessments of a wider range of stakeholders may get around this issue to some extent, but this may lead to difficulties in comparisons across local governments (or other geographical units). For example, if the range and characteristics of stakeholders varies substantially across geography, then their preferences, priorities, and ideologies may impact the types of assessments they provide for a given public service. A similar issue arises when using data from performance appraisals: the way that appraisals are implemented may vary across organizations and even across managers within the same organization, as management priorities, preferences, and practices vary.

## Micro: Independent Observers ('Mystery Shoppers') and Process Productivity

---

A more recent approach to measuring the performance of public officials involves the use of independent observers and standardized requests to compare different public-sector organizations. Such an approach involves requesting a standardized service from different providers and comparing the features of the response (speed, quality, equity, etc).

For example, researchers can send a standardized request to different public-sector agencies and compare how they deal with the request. This is the approach used in [Chong et al \(2014\)](#), where the authors send letters to non-existent addresses in 159 countries and measure the speed with which

the letters are returned to the US, providing a comparison of postal-service efficiency across countries. Another way to implement this measurement approach involves employing enumerators to use a public service in a standardized way and recording the quality of the service that they receive. Examples of this approach include using enumerators as standardized patients (presenting the same set of symptoms to different doctors) to measure the quality of care provided by healthcare workers ([Das and Hammer, 2007](#)) and using enumerators to request drivers' licenses and monitoring the speed of their delivery ([Bertrand et al, 2007](#)). These approaches allow for the performance of public officials to be measured while holding the type and complexity of the case, the identity of the user, and the framework behind the measure of satisfaction/subjective user assessments constant.

This method opens the door for capturing the performance for services and processes that have typically been more difficult to measure – those that do not necessarily have a final user. For example, productivity measures can be captured as long as there is a standardized request or an interaction between the requester and the public servant and a measure of performance (response time, quality of the interaction, whether they asked for the correct information, the number of repeated requests, etc) that can be recorded in a standardized way. This approach could, therefore, be well applied to middle-tier administrators.

However, an issue with this approach is whether the public official is aware that they are being monitored, which may have to be the case in some settings for feasibility and ethical reasons, and adjust their behaviours for the study.

This set of measures also relies on the existence of an identifiable set of good practices or measures that the independent observer can easily capture and record in a standardized way (e.g. response times, a checklist of correct questions asked, a checklist of information requested, a checklist of checks conducted, etc). As the assessments required by the observer become more complex, this approach can run into some of the difficulties associated with subjective assessments, described above.

## Micro: The Knowledge and Capacity of Public Officials

---

This approach aims to directly measure the capacity, knowledge, competencies, and/or skills of public officials. When used as a measure of performance or output, the implicit assumption behind this approach is that more knowledge is transferred into better performance, which may not be the

case, due to differences in the incentive environment and the effort exerted by public servants ([Das and Hammer, 2005](#)). An example of this approach is used in [Rogger and Somani \(2019\)](#), where civil servants are directly asked about their knowledge of fundamental features of their jurisdiction (e.g. the population, the unemployment rate, the level of rurality) and their sector of work (e.g. education officials asked about the enrolment rate, pupil-teacher ratio, etc). These estimates are then benchmarked against objective, official monitoring data. Another example of this approach involves asking clinicians a set of vignettes to measure their knowledge of best practices in dealing with a standardized set of cases ([Das and Hammer, 2005](#)). Another set of studies within this domain measures the accuracy of public officials' policy-effect estimates, willingness-to-pay for new information, and their use of new information in policy actions. This data allows researchers to assess officials' knowledge of their policy environment and the production function for public services, as well as the

rate at which they absorb and use new information ([Dal Bo et al. 2019](#); [Hjort et al. 2019](#)).

These methods produce micro-level data on specific margins of knowledge, skills, and competencies of public-sector employees. This data provides public-sector managers with direct operational information on where new training and capacity building efforts should focus, the effectiveness of previous training programs, as well as whether information (rather than, say, effort) is a bottleneck. However, information and capacity does not necessarily mean better performance. The performance of public officials may also depend on their available equipment, management, peers, and effort levels. This approach may also only be appropriate for certain public servants or certain public-sector tasks/processes, where the relationship between knowledge and output is clearer (such as in healthcare). Furthermore, some dimensions of knowledge, such as leadership quality, are very difficult to capture accurately.





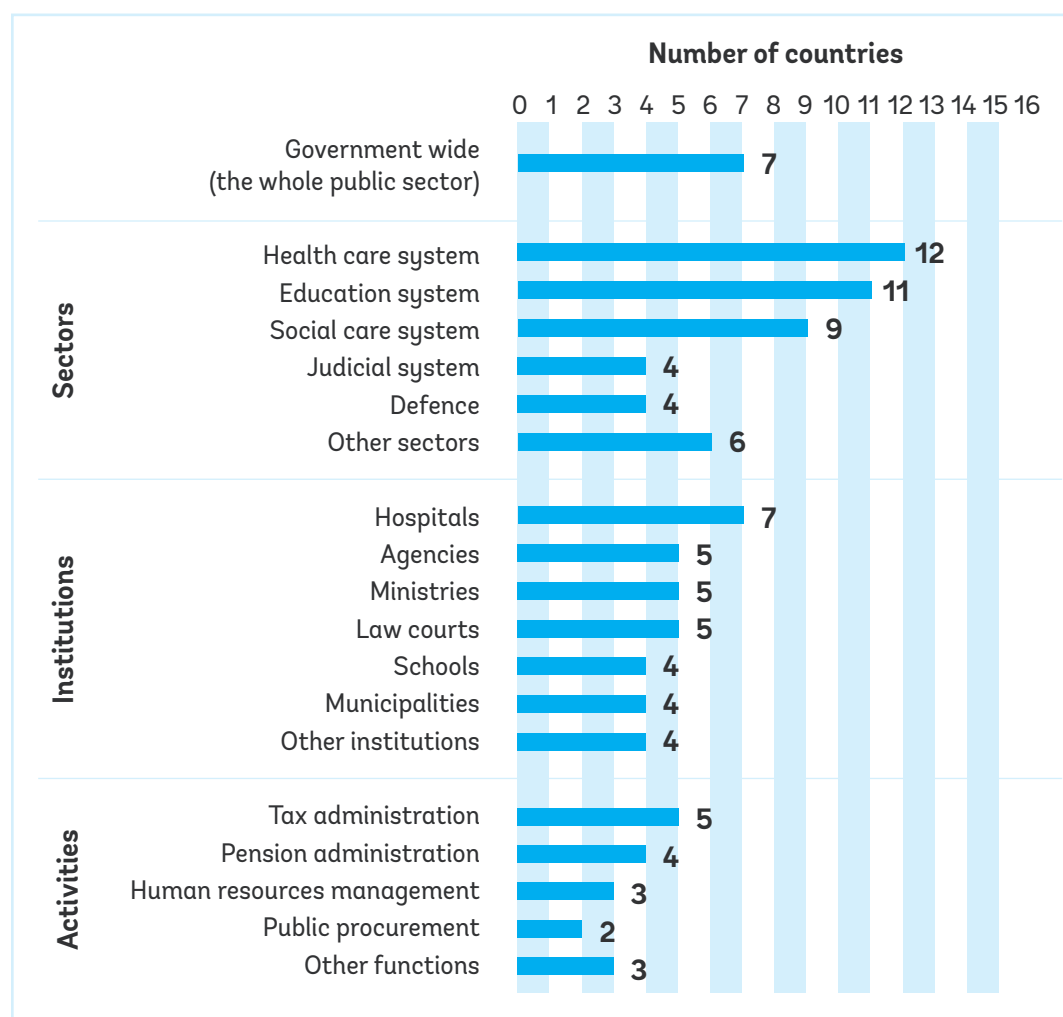
# Which Methods Are Commonly Used?

According to a recent study conducted by the OECD, only seven countries across the OECD report measures of the productivity of the whole public sector – Australia, Denmark, the Netherlands, New Zealand, Portugal, South Africa and the United Kingdom ([Lau et al., 2017](#)). Most focus only on individual sectors, particularly individual services, such as education and health; very few use productivity measures for individual institutions and activities.

This evidence implies that there is a lot of room for expanding measures of public-sector performance and productivity and embedding them into administrative functions. With the advent of more intensive micro-level data-collection efforts in the public sector and digital-government technologies, the opportunities for doing so are ever increasing.

> > >

**FIGURE 4 - Approaches to Measuring Public-Sector Productivity Across the OECD - Chart 2 from Lau et al (2017)**





# Conclusions and Recommendations

Accurately measuring public-sector productivity is important because of the size of the public sector in the market for goods, services, and labor. Achieving productivity growth in the public sector allows governments to produce more with less, an objective which has become especially important as a result of coronavirus-related pandemic and its impacts on public-sector finances.

This note summarizes the challenges of measuring productivity in the public sector, which revolve around the difficulty of accurately measuring public-sector outputs or linking outputs to specific inputs.

This note also outlines methods of directly measuring public-sector output, an alternative to the conventional approach of using inputs as measures of outputs. While many of these approaches provide new hopes for better measuring the productivity of the public sector, we have seen that all approaches have their own advantages and disadvantages—there is, unfortunately, no ‘silver bullet’! This section presents the main recommendations based on the review and a technical roadmap for implementing the measurement approaches described above.

## Recommendation 1. Combine Multiple Measures and Separately Report Inputs and Outputs (Performance)

---

**Given that there are strengths and weaknesses to each and every measurement approach, combining multiple measures of public-sector productivity will better capture a holistic picture of productivity and minimize the adverse incentive effects of focusing on a single measure.**

For example, using task-completion rates alone may shift the attention of public-sector managers towards performing only tasks that are easy and fast to complete and ignoring or not accepting cases that are more complex. In addition, the task-completion rate may increase, but certain immeasurable dimensions of quality may suffer, such as the long-term durability of constructed roads or water points. To minimize these adverse consequences, one can augment the use of task-completion rates with measures of user satisfaction and subjective assessments by a range of stakeholders in order to pick up early warning signs of deteriorating service quality.

Similarly, as discussed in the beginning of the note, the complex nature of the public-service production function can make it difficult to interpret changes in singular productivity metrics and the use of singular metrics may incentivize an excessive focus on cost-reduction at the expense of service quality.

For example, an increase in cost-weighted-output measures of productivity can arise from an increase in the quantity of outputs holding quality and costs constant, a decrease in input expenditure that may or may not manifest into future reductions in quality, or both of the above. When organizations are benchmarked and held accountable based on such measures and when quality is difficult to monitor or manifests several years after investments in inputs have been made, public-sector managers may prefer to actively reduce costs at the expense of service quality.

**Separately monitoring and reporting inputs expenditures and performance measures, including those attempting to capture service quality, allows for a better understanding of the driving factors behind changes in productivity.** Identifying *where* productivity changes are coming from is necessary to better understand the drivers of public-sector productivity and generate genuine long-run productivity growth run.

## Recommendation 2. Link Productivity Measures to the Service Delivery Chain and the Organization’s Production Function

To identify the most appropriate productivity measures, it may be helpful to conceptualize the service-delivery productivity chain and look for measures at each point in the

**chain.** As part of this exercise, it is also important to consider the production function (how each input interacts with one another) and the incentive environment within the unit or organization in question. Two examples are provided below, linked to two hypothetical public-sector-personnel reforms.

The first example concerns a training and capacity building program to implement a new best practice within the civil service. We might expect this reform to:

1. Improve the knowledge of civil servants about best practices
2. Improve individual measures of output or performance
  - Does the training and new set of best practices impact staff motivation?
  - Does the reform generate greater inequalities in performance?
3. Lead to improvements in service delivery, user satisfaction, and other assessments of organizational productivity
4. Is the reform cost-effective?

By drawing a chain from reform to desired outcomes, we can begin to identify the most appropriate measures of performance and their data sources to evaluate the effectiveness of the reform and identify any adverse consequences that may arise (e.g. lower employee motivation or user satisfaction). Figure 5 provides an example of such a chain and corresponding measures and sources of data.

> > >

**FIGURE 5 - An Example Reform Chain and Individual Measures of Outputs for a Training and Capacity Building Reform**



The second example concerns the introduction of an e-procurement system. We might expect this reform to:

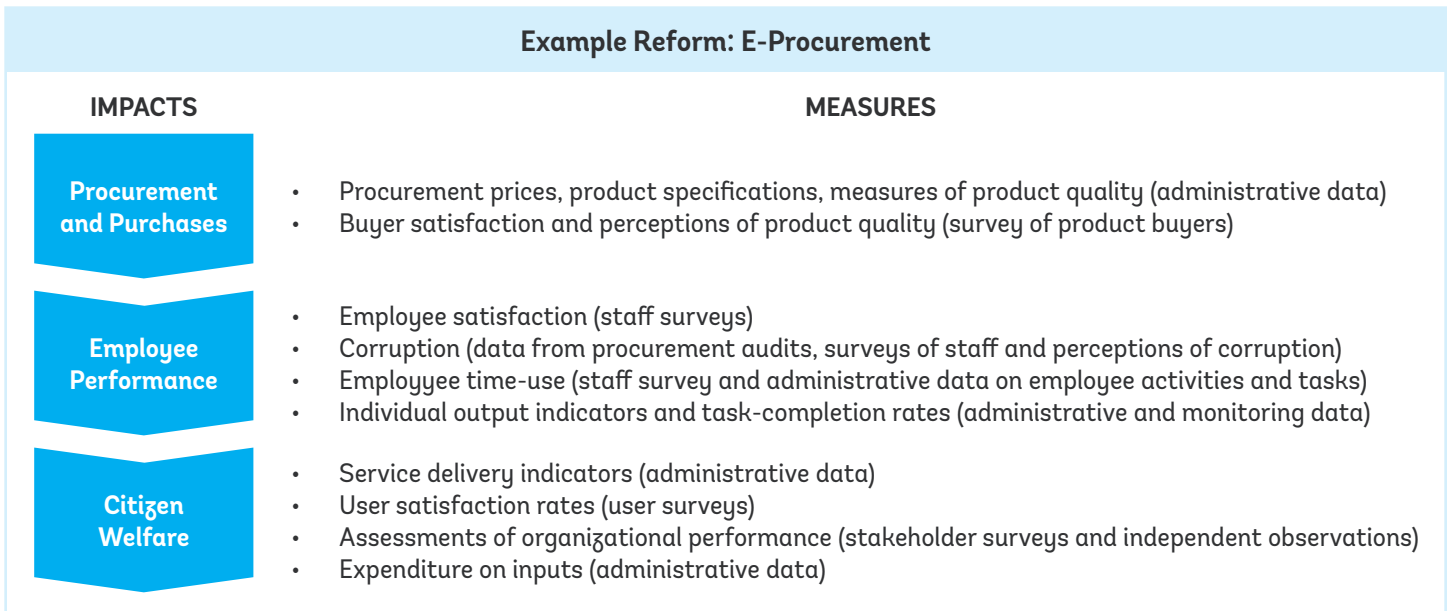
1. Reduce unit prices
  - Is there evidence of quality decreasing? Is there evidence that different types of products are purchased?
2. Affect employee time-use, behaviors, activities, and potentially satisfaction

3. Do employees have more time for other tasks; do they have more or fewer opportunities for corruption; are they less motivated for having to learn a new technology?
4. Impact measures of organizational performance and efficiency
  - Is the reform cost-effective?

Figure 6 presents these expected effects and their corresponding measures in a chain:

> > >

**FIGURE 6 - An Example Reform Chain and Individual Measures of Outputs for an E-Procurement Reform**



# A Technical Roadmap: Technical and Data Requirements for Implementation

Table 2, below, provides an outline of the technical and data requirements for implementing each of the measurement approaches described in this note.

> > >

**TABLE 2 - Technical and Data Requirements for Each Measurement Approach**

Macro (organization, sector, whole public service)	Micro (employee, task, process)
<p style="text-align: center;"><b>COST-WEIGHTED OUTPUT (<a href="#">Atkinson, 2005</a>)</b></p> <p><b>Technical requirements:</b></p> <ul style="list-style-type: none"> <li>• Data on organization’s outputs/activities for (e.g. no of licenses issued; no of audits; etc)</li> <li>• Cost data for each activity (total labor costs; equipment; intermediate inputs; etc)</li> <li>• Appropriate inflation measure that can be used to deflate costs over time</li> <li>• Measures (or proxies) of service quality</li> </ul>	<p style="text-align: center;"><b>PROCESS-, PROJECT- AND TASK-COMPLETION RATES</b></p> <p><b>Technical requirements:</b></p> <ul style="list-style-type: none"> <li>• Defining specific tasks, processes, and projects within the organization or sector</li> <li>• Data on start times, end times, and completion</li> <li>• Data on the process characteristics (e.g. complexity; quality, such as errors/complaints)</li> <li>• Assignment of individuals to tasks/processes allows employee-level analyses</li> <li>• Expenditure data for the tasks/projects/processes/organization and an appropriate inflation measure</li> </ul>
<p style="text-align: center;"><b>SERVICE DELIVERY INDICATORS</b></p> <p><b>Technical requirements:</b></p> <ul style="list-style-type: none"> <li>• Data on a range of indicators (e.g. enrolment rates, pupil-teacher ratios, test scores, etc)</li> <li>• Expenditure data for the sector (or specifically related to each indicator if possible) and an appropriate inflation measure</li> </ul>	<p style="text-align: center;"><b>PROCUREMENT OUTCOMES</b></p> <p><b>Technical requirements:</b></p> <ul style="list-style-type: none"> <li>• Data on prices (winning bid), quantity, and characteristics of items procured</li> <li>• Data on procuring officer (organization) allows analysis at the individual (organization) level</li> <li>• Data on procurement activities (date and duration of each process stage, number of bidders, quality of tender document) allows analysis of overall procurement process</li> </ul>
<p style="text-align: center;"><b>BUDGET EXECUTION RATES</b></p> <p><b>Technical requirements:</b></p> <ul style="list-style-type: none"> <li>• Budget data: plans and actual expenditure. Item-, project- or activity-level data allow for micro-level analyses</li> </ul>	<p style="text-align: center;"><b>STAFF AND USER SATISFACTION</b></p> <p><b>Technical requirements:</b></p> <ul style="list-style-type: none"> <li>• Surveys of staff/user satisfaction, using a comparable set of questions and scales</li> <li>• User surveys ideally conducted immediately after service (recall bias) and collect details of the request (e.g. complexity)</li> <li>• Representative sample of staff and users and high response rate avoids selection bias</li> <li>• Expenditure data for the service/organization and an appropriate measure of inflation</li> </ul>



### SUBJECTIVE ASSESSMENTS (BY EMPLOYEES, BY STAKEHOLDERS)

#### Technical requirements:

- Surveys of stakeholders on various dimensions of public-service performance using a comparable set of questions and scales
- Broad range of stakeholders (e.g. employees, civil society, media, managers, business associations, etc)
- Representative sample and high response rate avoids selection bias
- Expenditure data for the service/organization and an appropriate measure of inflation

### INDEPENDENT OBSERVERS (‘MYSTERY SHOPPERS’) AND PROCESS PRODUCTIVITY

#### Technical requirements:

- Identify a set of practices that can be independently verified and that signal quality (e.g. doctors asking the right questions for a given set of symptoms; officials making the necessary checks to an application; etc)
- Introduce independent observations of individuals/organizations to assess ‘business-as-usual’ responses to standardized requests
- Expenditure data for the service/organization and an appropriate measure of inflation

### THE KNOWLEDGE/CAPACITY OF PUBLIC OFFICIALS

#### Technical requirements:

- Comparable surveys and assessments of public officials’ information/knowledge
- Data on objective (correct/actual) practices or values to assess accuracy officials’ responses
- Representative sample of officials and high response rate avoids selection bias
- Expenditure data for the service/organization and an appropriate measure of inflation



