

The background features a series of curved lines of binary code (0s and 1s) in a light red color, receding into the distance. A solid dark red horizontal bar spans the width of the page, positioned above the main title.

PART 2

Foundational Themes in Government Analytics

CHAPTER 4

Measuring What Matters

Principles for a Balanced Data Suite That Prioritizes Problem Solving and Learning

Kate Bridges and Michael Woolcock

SUMMARY

Responding effectively and with professional integrity to public administration's many challenges requires recognizing that access to more and better quantitative data is necessary but insufficient. An overreliance on quantitative data comes with risks, of which public sector managers should be keenly aware. We focus on four such risks: first, that attaining easy-to-measure targets becomes a false standard of broader success; second, that measurement becomes conflated with what management is and does; third, that an emphasis on data inhibits a deeper understanding of the key policy problems and their constituent parts; and fourth, that political pressure to manipulate key indicators, if undetected, leads to falsification and unwarranted impact claims or, if exposed, jeopardizes the perceived integrity of many related (and otherwise worthy) measurement efforts. The cumulative concern is that these risks, if unattended to, will inhibit rather than promote public sector organizations' core problem-solving and implementation capabilities, an issue of high importance everywhere but especially in developing countries. We offer four cross-cutting principles for building an approach to the use of quantitative data—a "balanced data suite"—that strengthens problem solving and learning in public administration: (1) identify and manage the organizational capacity and power relations that shape data management; (2) focus quantitative measures of success on those aspects that are close to the problem; (3) embrace a role for qualitative data and a theory of change, especially for those aspects which require in-depth, context-specific knowledge; and (4) protect space for judgment, discretion, and deliberation because not everything that matters can be measured.

Kate Bridges is an independent consultant. Michael Woolcock is the lead social scientist in the World Bank's Development Research Department.

ANALYTICS IN PRACTICE

- Identify and manage the organizational capacity and power relations that shape data management. Make professional principles and standards for collecting, curating, analyzing, and interpreting data clear to all staff—from external consultants to senior managers—in order to affirm and enforce commitments to ensuring the integrity of the data themselves and the conclusions drawn from them. Make measurement accountable to advisory boards with relevant external members. Communicate measurement results to the public in a clear and compelling way, especially on contentious, complex issues.
- Focus quantitative measures of success on those aspects that are close to the problem. Ensure that the measurement approach itself is anchored to a specific performance problem. Target measurement investments at those performance problems that are prioritized by the administration. Ensure that any judgments on an intervention's success or failure are based on credible measures of the problem being fixed and not simply on output or process metrics. Where measures of success relate to whether the intervention is functioning, allow flexibility in the implementation of the intervention (where possible) and in the related measurement of its functioning. In this way, implementation strategies can shift if it becomes clear from the collected data that they are not making progress toward fixing the problem.
- Embrace an active role for qualitative data and a theory of change. Include qualitative data collection as a complement to quantitative data. This may be a prelude to future large-scale quantitative instruments or perhaps the only available data option for some aspects of public administration in some settings (such as those experiencing sustained violence or natural disasters). Draw on qualitative methods as a basis for eliciting novel or “unobserved” factors driving variation in outcomes. Tie measurement (both qualitative and quantitative) back to a theory of change. If the implementation of an intervention is not having its intended impact on the problem, assess whether there are mistaken assumptions regarding the theory of change.
- Protect space for judgment, discretion, and deliberation because not everything that matters can be measured. Consider carefully what you choose to measure, recognizing that whatever you choose will inevitably create incentives to neglect processes and outcomes that cannot be measured. Actively identify what you cannot (readily) measure that matters and take it seriously, developing strategies to manage that as well. Identify those aspects of implementation in the public sector that require inherently discretionary decisions. Employ strategies that value reasoned judgment and allow meaningful space for qualitative data inputs and the practical experience of embedded individuals, treating such inputs as having value alongside more quantitative ones.
- In the longer term, develop organizational systems that foster “navigation by judgment”—for example, a management structure that delegates high levels of discretion to allow those on the ground the space to navigate complex situations, recruitment strategies that foster high numbers of staff with extensive context-specific knowledge, and systems of monitoring and learning that encourage the routine evaluation of theory against practice.

INTRODUCTION

“What gets measured gets managed, and what gets measured gets done” is one of those ubiquitous (even clichéd) management phrases that hardly require explanation; it seems immediately obvious that the data generated by regular measurement and monitoring make possible the improvement of results. Less well known than the phrase itself is the fact that, although it is commonly attributed to the acclaimed management

theorist Peter Drucker, Drucker himself never actually said it (Zak 2013). In fact, Drucker's views on the subject were reportedly far more nuanced, along the lines of those of V. F. Ridgway, who argued over 65 years ago that not everything that matters can be measured and that not everything that can be measured matters (Ridgway 1956). Simon Caulkin (2008), a contemporary business management columnist, neatly summarizes Ridgway's argument, in the process expanding the truncated "to measure is to manage" phrase to "What gets measured gets managed—even when it's pointless to measure and manage it, and even if it harms the purpose of the organisation to do so."

Ridgway's and Caulkin's warnings—repeated in various guises by many since—remind us that the indiscriminate use of quantitative measures and undue confidence in what they can tell us may be highly problematic in certain situations, sometimes derailing the very performance improvements that data are intended to support (Merry, Davis, and Kingsbury 2015).¹ We hasten to add, of course, that seeking more and better quantitative data is a worthy aim in public administration (and elsewhere). Many important gains in human welfare (for example, recognizing and responding to learning disabilities) can be directly attributed to interventions conceived of and prioritized on the basis of empirical documentation of the reality, scale, and consequences of the underlying problem. The wonders of modern insurance are possible because actuaries can quantify all manner of risks over time, space, and groups. What we will argue in the following sections, however, is that access to quantitative data alone is not a sufficient condition for achieving many of the objectives that are central to public administration and economic development.

This chapter has five sections. Following this introduction, we lay out in section two how the collection, curation, analysis, and interpretation of data are embedded in contexts: no aspect takes place on a blank slate. On one hand, the institutional embeddedness of the data collection and usage cycle—in rich and poor countries alike—leaves subsequent delivery efforts susceptible to a host of possible compromises, stemming from an organization's inability to manage and deploy data in a consistently professional manner. At the same time, the task's inherent political and social embeddedness ensures it will be susceptible to influence by existing power dynamics and the normative expectations of those leading and conducting the work, especially when the political and financial stakes are high. In contexts where much of everyday life transpires in the informal sector—rendering it "illegible" to, or enabling it to actively avoid engagement with, most standard measurement tools deployed by public administrators—sole reliance on formal quantitative measures will inherently only capture a slice of the full picture.

In section 3, we highlight four specific ways in which an indiscriminate increase in the collection of what is thought to be "good data" can lead to unintended and unwanted (potentially even harmful) consequences. The risks are that (1) the easy-to-measure can become a misleading or false measure of broader reality, (2) measurement can become conflated with what management is and does, (3) an emphasis on what is readily quantified can inhibit a fuller and more accurate understanding of the underlying policy problem(s) and their constituent elements, and (4) political pressure to manipulate selected indicators, if undetected, can lead to falsification and unwarranted expectations—or, if exposed, can compromise the perceived integrity of otherwise worthy measurement endeavors.

Thankfully, there are ways to anticipate and mitigate these risks and their unintended consequences. Having flagged how unwanted outcomes can emerge, we proceed to highlight, in section 4, some practical ways in which public administrators might thoughtfully anticipate, identify, and guard against them. We discuss what a balanced suite of data tools might look like in public administration and suggest four principles that can help us apply these tools to the greatest effect, thereby enabling the important larger purposes of data to be served. For further methodological guidance, practitioners should consult appendix A, which provides a checklist titled "Using Expansive and Qualified Measurement for Informed Problem Solving and Learning in Public Administration." We stress from the outset that our concerns are not with methodological issues per se, or with the quality or comprehensiveness of quantitative data; these concerns are addressed elsewhere in *The Government Analytics Handbook* and in every econometrics textbook, and they should always be considered as part of doing "normal social science." The concerns we articulate are salient even in a best-case scenario, in which analysts have access to great data acquired from a robust methodology, although they are obviously compounded when the available data are of poor quality—as is often the case, especially in low-income countries—and when too much is asked of them.

HOW DATA ARE IMPACTED BY THE INSTITUTIONAL AND SOCIOPOLITICAL ENVIRONMENT IN WHICH THEY ARE COLLECTED

For all administrative tasks, but especially those entailing high-stakes decision-making, the collection and use of data is a human process inherently subject to human foibles (Porter 1995). This is widely accepted and understood: for example, key conceptual constructs in development (such as “exclusion,” “household,” and “fairness”) can mean different things to different people and translate awkwardly into different languages. With this in mind, professional data collectors will always give serious attention to “construct validity” concerns to ensure there is close alignment between the questions they ask and the questions their informants hear.² For present purposes, we draw attention to issues given less attention, but which are critical nonetheless—namely, the institutional and political factors that comprise the context shaping which data are (and are not) collected, how and from whom they are collected, how well they are curated over time, and how carefully conclusions and policy implications are drawn from analyses of them. We briefly address each item in turn.

INSTITUTIONAL EMBEDDEDNESS OF DATA

Beyond the purposes to which they are put, the careful collection, curation, analysis, and interpretation of public data are themselves complex technical and administrative tasks, requiring broad, deep, and sustained levels of organizational capability. In this section, we briefly explore three institutional considerations shaping these factors: the dynamics shaping the (limited) “supply” and refinement of technical skills, the forging of a professional culture that is a credible mediator of complex (and potentially heated) policy issues yet sufficiently robust to political pressure, and the related capacity to infer what even the best data analysis “means” for policy, practice, and problem solving.

These issues apply in every country but are especially salient in low-income countries, where the prevailing level of implementation capability in the public sector is likely to be low, and where the corresponding expectations of those seeking to improve it by expanding the collection and use of quantitative data may be high. At the individual level, staff with the requisite quantitative analytical skills are likely to be in short supply because acquiring such skills requires considerable training, while those who do have them are likely to be offered much higher pay in the private sector. (One could in principle outsource some data collection and analysis tasks to external consultants, but doing so would be enormously expensive and potentially compromise the integrity and privacy of unique public data.)

So understood, it would be unreasonable to expect the performance of data-centric public agencies to be superior to other service delivery agencies in the same context (for example, public health). Numerous studies suggest the prevailing levels of implementation capability in many (if not most) low-income countries are far from stellar (Andrews, Pritchett, and Woolcock 2017).³ For example, Jerven’s (2013) important work in Africa on the numerous challenges associated with maintaining the System of National Accounts—the longest-standing economic data collection task asked of all countries, from which their respective gross domestic products (GDPs) are determined—portends the difficulties facing less high-profile metrics (see also Sandefur and Glassman 2015).⁴ Put differently: if many developing countries struggle to curate the single longest-standing, universally endorsed, most important measure asked of them, on what basis do we expect these countries to manage lesser, lower-stakes measures?

To be sure, building quantitative analytical skills in public agencies is highly desirable; for present purposes, our initial point is a slight variation on the old adage that the quality of outcomes derived from quantitative data is only as good as the quality of the “raw material” and the competence with which it is analyzed and interpreted.⁵ Fulfilling an otherwise noble ambition to build a professional public sector whose decisions are informed by evidence requires a prior and companion effort to build the requisite

skills and sensibilities. Put differently, precisely because effective data management is itself such a complex and difficult task, in contexts where agencies struggle to implement even basic policy measures at a satisfactory level (for example, delivering mail and ensuring attendance at work), it is unlikely that, *ceteris paribus*, asking these agencies to also take a more “data-driven” approach will elicit substantive improvement. More and better “data” will not fix a problem if the absence of data is not itself the key problem or the “binding constraint”; the priority issue is discerning what *is* the key policy problem and its constituent elements. From this starting point, more and better data can be part of, but not a substitute for, strategies for enhancing the effectiveness of public sector agencies.

Even if both data management and broad institutional capability are functioning at high and complementary levels, there remains the structural necessity of interpreting what the data *mean*. Policy inference from even the best data and most rigorous methodology is never self-evident; it must always be undertaken in light of theory. This might sound like an abstract academic concern, but it is especially important when seeking to draw lessons from, or to make big decisions regarding the fate of, complex interventions. This is so because a defining characteristic of a complex problem is that it generates highly variable outcomes across time, space, and groups.

Promoting gender equality, for example, is a task that rarely generates rapid change: it can take a generation (or several, or centuries) for rules requiring equal participation in community meetings, or equal pay for equal work, to become the “new normal.”⁶ So, assessed over a five-year time frame, a “rigorous” methodology and detailed data may yield the empirical finding that a given gender empowerment project (GEP) has had “no impact”; taken at face value, this is precisely what “the data” would show and is the type of policy conclusion (“the GEP doesn’t work”) that would be drawn. However, interpreted in the light of a general theory of change incorporating the likely impact trajectory that GEP-type interventions follow—that is, a long period of stasis eventually leading to a gradual but sustained takeoff—a “doesn’t work” conclusion would be unwarranted; five years is simply too soon to draw a firm conclusion (Woolcock 2018).⁷ High-quality data and a sound methodology alone cannot solve this problem: a GEP may well be fabulous, neutral, useless, or a mixture of all three, but discerning which of these it is—and why, where, and for whom it functions in the way it does—will require the incorporation of different kinds of data into a close dialogue with a practical theory of change fitted for the sector, the context, and the development problem being addressed.

SOCIOPOLITICAL EMBEDDEDNESS OF DATA

Beyond these institutional concerns, a second important form of embeddedness shaping data collection, curation, and interpretation is the manner in which all three are shaped by sociopolitical processes and imperatives. All data are compiled for a purpose; in public administration, the scale and sophistication of the required data are costly and complex (requiring significant financial outlay and, thus, competition with rival claimants). Data are frequently called upon to adjudicate both the merits of policy proposals *ex ante* (for example, the Congressional Budget Office in the United States) and the effectiveness of programmatic achievements *ex post* (for example, the World Bank’s Independent Evaluation Group), which frequently entails entering into high-stakes political gambits—for example, achieving signature campaign proposals in the early days of an administration and proclaiming their subsequent widespread success (or failure) as election time beckons again. (See more on this below.)

Beyond the intense political pressure “data” are asked to bear in such situations, a broader institutional consideration is the role large-scale numerical information plays in “rendering legible” (Scott 1998) complex and inherently heterogeneous realities, such that they can be managed, mapped, and manipulated for explicit policy purposes. We hasten to add that such “thin simplifications” (Scott’s term) of reality can be both benign and widely beneficial: comprehensive health insurance programs and pension systems have largely tamed the otherwise debilitating historical risks of, respectively, disease and old age by generating premiums based on

general demographic characteristics and the likelihood of experiencing different kinds of risks (for example, injuries or cancer) over the course of one's life.

A less happy aspect of apprehending deep contextual variation via simplified (often categorical) data, however, is the corresponding shift it can generate in the political status and salience of social groups. The deployment of the census in colonial India, for example, is one graphic demonstration of how the very act of “counting” certain social characteristics—such as the incidence of caste, ethnicity, and religion—can end up changing these characteristics themselves, rendering what had heretofore been relatively fluid and continuous categories as fixed and discrete. In the case of India, this massive exercise in data collection on identity led to “caste” being created, targeted, and mobilized as a politically salient characteristic that had (and continues to have) deep repercussions (for example, at independence, when Pakistan split from India, and more recently within the rise of Hindu nationalism; see Dirks 2011).⁸ Similarly, influential scholars have argued that the infamous Hutu/Tutsi massacre in Rwanda was possible at the scale at which it was enacted because ethnic categories were formalized and fixed via public documents whose origins lie in colonial rule (for example, Mamdani 2002).

For Scott (1998), public administration can only function to the extent its measurement tools successfully turn widespread anthropological variation, such as in languages spoken, into singular modern categories and policy responses—for instance, to ensure that education is conducted in one national language, in a school, and on the basis of a single curriculum.⁹ The net welfare gains to society might be unambiguous, but poorer, isolated, marginalized, and less numerous groups are likely to bear disproportionately the costs of this trade-off. If official “data” themselves constitute an alien or distrusted medium by which certain citizens are asked to discern the performance of public agencies, merely providing (or requiring) “more” is unlikely to bring about positive change. In such circumstances, much antecedent work may need to be undertaken to earn the trust of citizens and to help them more confidently engage with their administrative systems.¹⁰ By way of reciprocity, it may also require such systems to interact with citizens themselves in ways that more readily comport with citizens' own everyday (but probably rather different) vernacular for apprehending the world and interpreting and responding to events. Either way, it is critical that officials be wary of the potentially negative or unintended effects of data collection, even when it may begin with a benign intention to facilitate social inclusion and more equitable policy “targeting.”¹¹

THE UNINTENDED CONSEQUENCES OF AN INDISCRIMINATE PURSUIT OF “MORE DATA”

There is a sense in which it is axiomatic that more and better data are always a good thing. But the institutional and sociopolitical embeddedness of data generation and the use of data in public administration (as discussed in the preceding section) means we need to qualify this assertion by focusing on where and how challenges arise. With this in mind, we turn our attention to instances where the increased collection of what is thought to be “good data” leads to perverse outcomes. Here, we highlight four such outcomes that may materialize as the result of an undue focus on issues, concepts, inputs, or outcomes that happen to be most amenable to being quantified.

Outcome 1: The Easy-to-Measure Becomes a False Standard of Success

What may start as a well-intentioned managerial effort to better quantify meaningful success can instead generate a blinkered emphasis on that which is simply easiest to quantify. The result can be a skewed or false sense of what a project has (or has not) achieved, and how, where, and for whom outcomes have been achieved.

In a recent study, we demonstrate how a variety of institutional incentives align across the government of Malawi and the World Bank in such a way that both government and World Bank officials consistently favor

easy-to-measure indicators (inputs and outputs, or what we refer to as “changes in form rather than function”) as the yardstick of project success (Bridges and Woolcock 2017). This is a quintessential example of what strategy writer Igor Ansoff describes as a situation in which “managers start off trying to manage what they want, and finish up wanting what they can measure” (quoted in Cahill 2017, 152). As a result of evaluating public financial management (PFM) projects that were implemented over the course of 20 years in Malawi, we show that almost 70 percent of what projects measure or aim for is “change in terms of whether institutions look like their functioning counterparts (that is, have the requisite structures, policies, systems, and laws in place),” whereas only 30 percent of what is measured can be said to be “functional”—that is, focused on “purposeful changes to budget institutions aimed at improving their quality and outcomes” (Andrews 2013, 7). What is more, we find that World Bank PFM projects have considerably more success in achieving “formal” results than “functional” ones. Unsurprisingly, demonstrable improvement in actual performance is far harder to achieve than change that is primarily regulative, procedural, or systems oriented. Unfortunately, an emphasis on what is easy-to-measure obfuscates this reality and allows reform “success” to be claimed.

In practice, Malawi’s history of PFM reform is littered with projects that claim “success” based on hardware procured, software installed, legislation developed, and people trained, whereas even a basic analysis reveals stagnation or even regression in terms of more affordable spending decisions, spending that reflects budgeted promises, greater ability to track the flow of funds, or reduction in corruption. As long as the World Bank and the Malawian government focus on “formal” measures, they are able to maintain the illusion of success: that is, until something like Malawi’s 2013 “Cashgate” crisis—in which it was revealed that about US\$32 million in government funds had been misappropriated between April and September 2013—lifts the lid on the deep-rooted financial management problems that have remained largely unaffected by millions of dollars of reform efforts. In this sense, Malawi is a microcosm of many institutional reform efforts globally. Although similar financial reforms have been globally implemented in a manner that suggests some level of consensus about “what works,” the outcomes of those reforms are varied at best and often considerably lower than anticipated (Andrews 2013).

In the same way that an emphasis on the easy-to-measure can lead to overestimation of success, it can also contribute to underestimation. Reforms can sometimes yield meaningful change via what McDonnell (2017) calls “the animating spirit of daily practice” but end up being missed because managers do not have good means of measuring, attributing, and enhancing these *kinds* of shifts. For example, when researching the impact of technical assistance on a large government health program in Nigeria, we found that there were strong indications that important innovations and shifts took place at the local level, including in aspects as difficult to shift as cultural practices regarding contraceptives (Bridges and Woolcock 2019). These shifts in practice and their impact on contraceptive uptake could not be apprehended by aggregated statewide indicators, however, and since no measurement was being done below this level, the progress and valuable lessons of such interventions were being missed.

Another example of the importance of having access to a broader suite of data comes from an assessment of a program in rural India seeking to promote participatory democracy in poor communities, where the curation of such a data suite enabled more nuanced and constructive lessons to be drawn (see Rao, Ananthpur, and Malik 2017). The results of the initial randomized controlled trial (RCT) deemed the program to have had no mean impact—and if these were the only data available, that would have been the sole conclusion reached.¹² Upon closer inspection, however, it was learned that there was considerable variation in the program’s impact. The average of this variation may have been close to zero, but for certain groups, the program had worked quite well, for others it had had no impact, while for still others it had been detrimental. Who were these different groups, and what was it about them that led to such variable outcomes? A companion qualitative process evaluation was able to discern that the key differences were the quality of implementation received by different groups, the level of support provided to them by managers and political leaders, and variations in the nature and extent of local-level inequalities (which, in turn, shaped which groups were able to participate and on what terms).¹³ The administrative rules and implementation guidelines provided to all groups were identical, but in this case, a qualitative process evaluation was able to document the ways

and places in which variable fidelity to them yielded widely different outcomes (albeit with no net impact). Moreover, the qualitative data were able to discern subtle positive effects from the program that the quantitative survey instrument alone would have missed.

Outcome 2: Measurement Becomes Conflated with Management

An extension of the above point is that an undue emphasis on quantitative data can lead measurement to become a *substitute for* rather than a *complement to* management. This is evident when only that which is quantifiable receives any significant form of managerial attention, an outcome made possible when the easily quantifiable becomes the measure of success, becoming, in turn, the object of management's focus, typically to the exclusion of all else. As Wilson (1989, 161) famously intoned in a classic study of bureaucratic life, "Work that produces measurable outcomes tends to drive out work that produces immeasurable outcomes."

In one sense this is hardly surprising: the need for managers to make decisions on the basis of partial information is difficult and feels risky, so anything that claims to fill that gap and bypass the perceived uncertainty of subjective judgment will be readily welcomed. "The result," Simon Caulkin (2016) argues, "both practically and theoretically, is to turn today's management into a technology of control that attempts to minimise rather than capitalise on the pesky human element." And in public administration, a managing-it-by-measuring-it bias can mean that, over time, the bulk of organizational resources end up neglecting the "pesky human element" of change processes, even though it is this element that is often central to attaining the transformational outcomes managers are seeking.

This dynamic characterizes key aspects of the Saving One Million Lives (SOML) initiative, an ambitious health sector reform program launched by the government of Nigeria. The original goal of SOML was to save the lives of one million mothers and children by 2015; to this end, SOML gave priority to a package of health interventions known as "the six pillars."¹⁴ The World Bank actively supported SOML, using its Program-for-Results (PforR) instrument to reward Nigerian states financially based on improvements from their previous best performance on the six key indicators.¹⁵ Improvements were to be measured through yearly household surveys providing robust estimates at the state level.

In practice, of course, these six pillars (or intervention areas) were wildly different in their drivers and complexity; improvement within them was therefore destined to move at different trajectories and different speeds for different groups in different places. State actors, keen to raise their aggregate measure of success and get paid for it, soon realized that there was some gaming to be done. Our field research documents how the emphasis on singular measures of success introduced a perverse incentive for states to focus on the easier metrics at the expense of the harder ones (Bridges and Woolcock 2019). Interviews with state officials revealed that frontline staff increasingly focused their time and energies on those constituent variables that they discerned were easiest to accomplish (for example, dispensing vitamin supplements) over those that were harder or slower—typically those that involved a plethora of "pesky human elements," such as lowering maternal mortality or increasing contraceptive use. In selecting certain outcomes for measurement and managing these alone, others inevitably end up being sidelined.

Likewise, a recent report on results-based financing (RBF) in the education sector (Dom et al. 2020) finds evidence of a "diversion risk" associated with the signposting effect of certain reward indicators, with important areas deprioritized because of the RBF incentive. For example, in Mozambique, they find that an emphasis on simple process indicators and a focus on targets appears to have led officials to divert resources and attention away from "more fundamental and complex issues," such as power dynamics in the school council, the political appointment of school directors, and teachers' use of training. Dom et al. also report evidence of "cherry-picking risks," in which less costly or politically favored subgroups or regions see greater resources, in part because they are more likely to reach a target. For example, in Tanzania, they find evidence that the implementation of school rankings based on exam results was correlated with weaker students not sitting, presumably in an effort by the schools to raise average exam pass rates.

This tendency becomes a particular issue when the sidelined outcomes end up being the ones we care most about. Andrew Natsios (2011), the former administrator of the United States Agency for International

Development (USAID, an organization charged with “demonstrating the impact of every aid cent that Congress approves”), argues compellingly that the tendency in aid and development toward what he calls “obsessive measurement disorder” is a manifestation of a core dictum among field-based development practitioners—namely, “that those development programs that are most precisely and easily measured are the least transformational, and those programs that are most transformational are the least measurable.” The change we often desire most is in very difficult-to-measure aspects, such as people’s habits, cultural norms, leadership characteristics, and mindsets.

This reality is also aptly illustrated in many anticorruption efforts, where imported solutions have managed to change the easy-to-measure—new legislation approved, more cases brought, new financial systems installed, more training sessions held—but have failed to shift cultural norms regarding the unacceptability of whistleblowing or the social pressures for nepotism (Andrews 2013). Failure to measure and therefore manage these informal drivers of the problem ensures that any apparent reduction in fund abuses tends to be short-lived or illusory.

This phenomenon is hardly limited to poor countries. A more brutal example of how what cannot be measured does not get managed, with disastrous results, can be found in the United Kingdom’s National Health System (NHS). While investigating the effects of competition in the NHS, Propper, Burgess, and Gossage (2008) discovered that the introduction of interhospital competition improved waiting times while also substantially *increasing* the death rate following emergency heart attacks. The reason for this was that waiting times were being measured (and therefore managed), while emergency heart-attack deaths were not tracked and were thus neglected by management. The result was shorter waiting times but more deaths as a result of the choice of measure. The authors note that the issue here was not intent but the extent to which one target consumed managerial attention to the detriment of all else; as they note, it “seems unlikely that hospitals deliberately set out to decrease survival rates. What is more likely is that in response to competitive pressures on costs, hospitals cut services that affected [heart-attack] mortality rates, which were unobserved, in order to increase other activities which buyers could better observe” (Propper, Burgess, and Gossage 2008).

More recently, in October 2019, the Global Health Security Index sought to assess which countries were “most prepared” for a pandemic, using a model that gave the highest ranking to the United States and the United Kingdom, largely on the basis of these countries’ venerable medical expertise and technical infrastructure, factors which are readily measurable (McCarthy 2019). Alas, the model did not fare so well when an actual pandemic arrived soon thereafter: a subsequent analysis, published in *The Lancet* on the basis of pandemic data from 177 countries between January 2020 and September 2022, found that “pandemic-preparedness indices . . . were not meaningfully associated with standardised infection rates or IFRs [infection/fatality ratios]. Measures of trust in the government and interpersonal trust, as well as less government corruption, had larger, statistically significant associations with lower standardised infection rates” (Bollyky et al. 2022, 1).

Needless to say, variables such as “trust” and “government corruption” are hard to measure, are hard to incorporate into a single theory anticipating or informing a response to a pandemic, and map awkwardly onto any corresponding policy instrument. For present purposes, the inference we draw from these findings is not that global indexes have no place; rather, they suggest the need, from the outset, for curating a broad suite of data when anticipating and responding to complex policy challenges, the better to promote real-time learning. Doubling down on what can be readily measured limits the space for eliciting those “unobserved” (and perhaps unobservable) factors that may turn out to be deeply consequential.

Outcome 3: An Undue Emphasis on Data Inhibits Understanding of the Foundational Problem(s)

An indiscriminate emphasis on aggregated, quantitative data can erode important nuances about the root causes of the problems we want to fix, thereby hampering our ability to craft appropriate solutions and

undermining the longer-term problem-solving capabilities of an organization. All too often, the designation of indicators and targets has the effect of causing people to become highly simplistic about the problems they are trying to address. In such circumstances, what should be organizational meetings held to promote learning and reflection on what is working and what is not instead become efforts in accounting and compliance (Honig and Pritchett 2019). Reporting, rather than learning, is incentivized, and management increasingly focuses on meeting target numbers rather than solving the problem. Our concern here is that, over time, this tendency progressively erodes an organization's problem-solving capabilities.

The education sector is perhaps the best illustration of this: time and again practitioners have sought to codify “learning,” and time and again this has resulted in an obfuscation of the actual causes underlying the problem. In a well-intentioned effort to raise academic performance, “the standards movement” in education promoted efforts hinged on quantitative measurement, as reported in the league tables of the Program for International Student Assessment (PISA).¹⁶ PISA runs tests in mathematics, reading, and science every three years with groups of 15-year-olds in countries around the world. Testing on such a scale requires a level of simplicity and “standardization,” thus the emphasis is on written examinations and the extensive use of multiple-choice tests so that students’ answers can be easily codified and processed (Robinson and Aronica 2015). Demonstrating competence in fundamental learning tasks certainly has its place, but critics have increasingly argued that such tests are based on the incorrect assumption that what drives successful career and life outcomes is the kind of learning that is capable of being codified via a standardized test (Claxton and Lucas 2015; Khan 2012).

In reality, the gap between the skills that children learn and are tested for and the skills that they need to excel in the 21st century is becoming more obvious. The World Economic Forum noted in 2016 that the traditional learning captured by standardized tests falls short of equipping students with the knowledge they need to thrive.¹⁷ Yong Zhao (2012), the presidential chair and director of the Institute for Global and Online Education in the College of Education at the University of Oregon, points out that there is an inverse relationship between those countries that excel in PISA tests and those that excel in aspects like entrepreneurship, for example (see figure 4.1).

While a focus on assessing learning is laudable—and a vast improvement over past practices (for example, in the Millennium Development Goals) of merely measuring attendance (World Bank 2018)—for present purposes the issue is that the drivers of learning outcomes are far more complex than a quantifiable content deficit in a set of subjects. This is increasingly the case in the 21st century, which has brought a need for new skills and mindsets that go well beyond the foundational numeracy and literacy skills required during the Industrial Revolution (Robinson and Aronica 2015). A survey of chief human resources and strategy officers by the World Economic Forum (2016) finds a significant shift between 2015 and 2020 in the top skills future workers will need, with “habits of mind” like critical thinking, creativity, emotional intelligence, and problem solving ranking well ahead of any specific content acquisition. None of this is to say that data do not have a role to play in measuring the success of an educational endeavor. Rather, the data task in this case needs to be informed by the complexity of the problem and the extent to which holistic learning resists easy quantification.¹⁸

Finally, relying exclusively on high-level aggregate data can result in presuming uniformity in underlying problems and thus lead to the promotion of simplistic and correspondingly generic solutions. McDonnell (2020) notes, for example, that because many developing countries have relatively high corruption scores, an unwelcome outcome has been that *all* the institutions in these countries tend to be regarded by would-be reformers as similarly corrupt and uniformly ineffectual. In her impressive research on “clusters of effectiveness,” however, she offers evidence of the variation in public-sector performance within states, noting how the aggregated data on “corruption” masks the fact that the difference in corruption scores between Ghana’s best- and worst-rated state agencies approximates the difference between Belgium (WGI = 1.50) and Mozambique (WGI = -0.396), in effect “spanning the chasm of so-called developed and developing worlds.” The tendency of reform actors to be guided by simplistic aggregate indicators—such as those that are used to determine a poor country’s “fragility” status and eligibility for International Development Association (IDA) funding—has prevented a more

FIGURE 4.1 Country Scores on Program for International Student Assessment Tests and Perceived Entrepreneurial Capability



Source: Based on Zhao 2012.
 Note: PISA = Program for International Student Assessment.

detailed and context-specific understanding of the lessons that could be drawn from positive outlier cases, or what McDonnell refers to as “the thousand small revolutions quietly blooming in rugged and unruly meadows.”¹⁹

Outcome 4: Pressure to Manipulate Key Indicators Leads to Falsification and Unwarranted Impact Claims

As an extension of our previous point regarding how the easy-to-measure can become the yardstick for success, it is important to acknowledge that public officials are often under extreme pressure to demonstrate success in selected indicators. Once data themselves, rather than the more complex underlying reality, become the primary objective by which governments publicly assess (and manage) their “progress,” it is inevitable that vast political pressure will be placed on these numbers to bring them into alignment with expectations, imperatives, and interests. Similar logic can be expected at lower units of analysis (for example, field offices), where it tends to be even more straightforward to manipulate data entry and analysis. This, in turn, contributes to a perverse incentive to falsify or skew data, to aggregate numbers across wildly different variables into single indexes, and to draw unwarranted inferences from them.

This risk is particularly acute, for instance, when annual global rankings are publicly released (assessing, for example, a country’s “investment climate,” “governance,” and gender equity), thereby shaping major investment decisions, credit ratings, eligibility for funding from international agencies, and the fate of senior officials charged with “improving” their country’s place in these global league tables. Readers will surely be aware of the case at the World Bank in September 2021, when an external review revealed that the *Doing Business* indicators had been subject to such pressure, with alterations made to certain indicators from certain countries (WilmerHale 2021). Such rankings are now

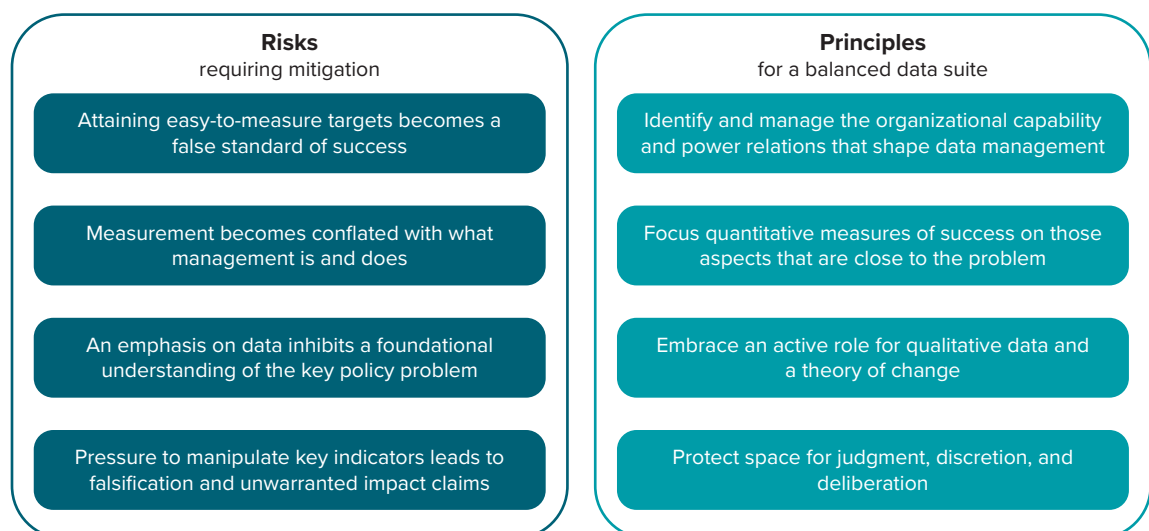
omnipresent, and if they are not done by one organization, they will inevitably be done by another. Even so, as the *Economist* (2021) magazine concluded, some might regard the *Doing Business* episode as “proof of ‘Goodhart’s law,’ which states that when a measure becomes a target, it ceases to be a good measure.” At the same time, it pointed out that there is a delicate dance to be done here, since “the *Doing Business* rankings were always intended to motivate as well as measure, to change the world, not merely describe it,” and “if these rankings had never captured the imagination of world leaders, if they had remained an obscure technical exercise, they might have been better as measures of red tape. But they would have been worse at cutting it.”

Such are the wrenching trade-offs at stake in such exercises, and astute public administrators need to engage in them with their eyes wide open. Even (or especially) at lower units of analysis, where there are perhaps fewer prying eyes or quality-control checks, the potential is rife for undue influence to be exerted on data used for political and budgetary-allocation purposes. Fully protecting the integrity of data collection, collation, and curation (in all its forms) should be a first-order priority, but so, too, is the need for deploying what should be standard “risk diversification” strategies on the part of managers—namely, not relying on single numbers or methods to assess inherently complex realities.

PRINCIPLES FOR AN EXPANSIVE, QUALIFIED DATA SUITE THAT FOSTERS PROBLEM SOLVING AND ORGANIZATIONAL LEARNING

In response to the four risks identified above, we offer a corresponding set of cross-cutting principles for addressing them. Figure 4.2 summarizes the four risks in the left-hand column and presents the principles as vertical text on the right, illustrating the extent to which the principles, when applied in combination, can serve to produce a more balanced data suite that prioritizes problem solving and learning.

FIGURE 4.2 Four Risks with Corresponding Principles for Mitigating Them to Ensure a Balanced Data Suite



Source: Original figure for this publication.

Note: The principles are “cross-cutting,” in the sense that they apply in some measure to all the risks; they are not one-to-one.

Principle 1: Identify and Manage the Organizational Capacity and Power Relations That Shape Data Management

The data collection and curation process takes place not in isolation but in a densely populated political and institutional ecosystem. It is difficult, expensive, and fraught work; building a professional team capable of reliably and consistently doing this work—from field-level collection and curation at headquarters to technical analysis and policy interpretation—will be as challenging as it is in every other public sector organization. Breakdowns can happen at any point, potentially compromising the integrity of the entire endeavor. For this reason, it is important for managers not just to hire those with the requisite skills but to cultivate, recognize, and reward a professional ethos wherein staff members can do their work in good faith, shielded from political pressure. Such practices, in turn, need to be protected by clear, open, and safe procedures for staff to report undue pressure, complemented by accountability to oversight or advisory boards comprising external members selected for their technical expertise and professional integrity. In the absence of such mechanisms, noble aspirations for pursuing an “evidence-based policy” agenda risk being perceived as means of providing merely “policy-based evidence.”

The contexts within or from which data are collected are also likely to be infused with their own socio-political characteristics. Collecting data on the incidence of crime and violence, for example, requires police to faithfully record such matters and their response to them, but they must do so in an environment where there may be strong pressure to underreport, whether because of personal safety concerns, lack of administrative resources, or pressure to show that a given unit’s performance is improving (where this is measured by showing a “lower incidence” of crime). In this respect, good diagnostic work will reveal the contours of the institutional and political ecosystem wherein the data work will be conducted and the necessary authorization, financing, and protection sought; it will also help managers learn how to understand and successfully navigate this space.²⁰ The inherent challenges of engaging with such issues might be eased somewhat if those closest to them see data deployment not as an end in itself or an instrument of compliance but as a means to higher ends—namely, learning, practical problem solving, and enhancing the quality of policy options, choices, and implementation capability.²¹

A related point is that corresponding efforts need to be made to clearly and accurately communicate to the general public those findings that are derived from data managed by public administrators, especially when these findings are contentious or speak to inherently complex issues. This has been readily apparent during the COVID-19 pandemic, with high-stakes policy decisions (for example, requiring vulnerable populations to forgo income) needing to be made on the basis of limited but evolving evidence. Countries such as Vietnam have been praised for the clear and consistent manner in which they issued COVID-19 response guidelines to citizens (Ravallion 2020), but the broader point is that even when the most supported decisions are based on the best evidence generated by the most committed work environments, it remains important for administrators to appreciate that the very act of large-scale measurement and empirical interpretation, especially when enacted by large public organizations, can potentially be threatening to or misunderstood by the very populations they are seeking to assist.

Principle 2: Focus Quantitative Measures of Success on Those Aspects That Are Close to the Problem

If we wish to guard against the tendency to falsely ascribe success based on the achievement of poorly selected indicators, then we should ensure that any indicators used to claim or deny reform success are as readily operational and close to the service delivery problem as possible. Output and process indicators are useful in their own ways, but we should not make the mistake of conflating their achievement with “problem fixed.” There are often strong institutional incentives to claim reform success based on whether a new mechanism or oversight structure has been created, a new law has been passed, or a percentage of participation has been achieved, but if meaningful change is sought, these incentives need to be countered. All of these

measures are changes in *form* that, while useful as indicators of outputs being met, can be achieved (and have been in the past) without any attendant functional shifts in the underlying quality of service delivery.

Officials can guard against this tendency by taking time to ensure that an intervention is focused on specific problems, including those that matter at a local level, and that the intervention's success and its attendant metrics are accurate measures of the problems being fixed. Tools like the Problem-Driven Iterative Adaptation (PDIA) Toolkit, designed by members of the Building State Capability program at the Center for International Development (CID) at Harvard University, can help guide practitioners in this process.²² The PDIA approach is designed to help practitioners break down their problems into root causes, identify entry points, search for possible solutions, take action, reflect upon what they have learned, adapt, and then act again. By embedding any intervention in such a framework, practitioners can ensure that success metrics are well linked to functional responses to locally felt problems.

Whatever tool is applied, the goal should be to arrive at metrics of success that represent a compelling picture of the root performance problem being addressed (and hopefully solved). So, for example, in our education example, metrics such as the number of teachers hired, the percentage of the budget dedicated to education, and the number of schools built are all output measures that say nothing about actual learning. Of course, there are assumptions that these outputs *lead* to children's learning, but as many recent studies now show, such assumptions are routinely mistaken; these indicators can be achieved even as actual learning regresses (Pritchett 2013; World Bank 2018). By contrast, when a robust measure of learning—in this case, literacy acquisition—was applied in India, it allowed implementers to gain valuable insights about which interventions actually made a difference, revealing that teaching to a child's actual level of learning, not their age or grade, led to marked and sustained improvements. Crucially, such outcomes are the result of carefully integrated qualitative and quantitative approaches to measurement (Banerjee et al. 2016).

Going further, various cross-national assessments around the world are trying to tackle the complex challenge of finding indicators that measure learning not just in the acquisition of numeracy, science, and literacy skills but in competencies that are increasingly valuable in the 21st century: grit, curiosity, communication, leadership, and compassion. PISA, for example, has included an “innovative domain” in each of its recent rounds, including creative problem solving in 2012, collaborative problem solving in 2015, and global competence in 2018. In Latin America, the Latin American Laboratory for Assessment of the Quality of Education (LLECE) included a module on socioemotional skills for the first time in its assessment of sixth-grade students in 2019, focusing on the concepts of conscience, valuing others, self-regulation, and self-management (Global Partnership for Education 2020). Much tinkering remains to be done, but the increase in assessments that include skills and competencies such as citizenship (local and global), socioemotional skills, information and communication technology literacy, and problem solving is a clear indication of willingness to have functional measures of success, capturing outcomes that matter.

In summary, then, public administrators who wish to guard against unwarranted impact claims and ensure metrics of success are credible can begin by making sure that an intervention itself is focused on a specific performance problem that is locally prioritized and thereafter ensure that any judgment of that intervention's success or failure is based not on output or process metrics but on measures of the problems being fixed. And having ensured that measures of success are functional, practitioners must allow for flexibility of implementation where possible so strategies can shift if it becomes clear from the collected data that they are not making progress toward fixing the problem, possibly due to mistaken assumptions regarding their theory of change.

Principle 3: Embrace an Active Role for Qualitative Data and a Theory of Change

The issues we have raised thus far, we argue, imply that public administrators should adopt a far more expansive concept of what constitutes “good data”—namely, one that includes insights from theory and qualitative research. Apprehending complex problems requires different forms and sources of data; correctly interpreting empirical findings requires active dialogue with reasoned expectations about what outcomes should be attained by when. Doing so helps avoid creating distortions that can generate (potentially wildly) misleading claims regarding “what's going on and why” and “what should be done.”

Specifically, we advocate for the adoption of a complementary *suite* of data forms and sources that favors flexibility, is focused on problem solving (as opposed to being an end in itself), and values insights derived from seasoned experience. In the examples we have explored above, reliance on a single form of data (sometimes even a single number) rendered projects vulnerable to political manipulation, unwarranted conclusions, and an inability to bear the decision-making burdens thrust upon them. More constructively, it was the incorporation of alternative methods and data in dialogue with a reasoned theory of change that enabled decision-makers to anticipate and address many of these same concerns.

To this end, we have sought to get beyond the familiar presumption that the primary role of qualitative data and methods in public administration research (and elsewhere) is to provide distinctive insights into the idiosyncrasies of an organization's "context" and "culture" (and thus infuse some "color" and "anecdotes" for accompanying boxes).²³ Qualitative approaches can potentially yield unique and useful material that contributes to claims about *whether* policy goals are being met and delivery processes duly upheld (Cartwright 2017); they can be especially helpful when the realization of policy goals requires integrating both adaptive and technical approaches to implementation—for example, responding to COVID-19. But perhaps the more salient contributions of qualitative approaches, we suggest, are to explore *how*, *for whom*, and *from whom* data of all kinds are deployed as part of broader imperatives to meet political requirements and administrative logics in a professional manner and to elicit novel or previously "unobserved" variables shaping policy outcomes.

Principle 4: Protect Space for Judgment, Discretion, and Deliberation

Our caution is against using data reductively: as a replacement or substitute for managing. Management *must* be about more than measuring. A good manager needs to be able to accommodate the immeasurable because so much that is important to human thriving is in this category; dashboards etc. certainly have their place, but if these were all that was needed, then "managing" could be conducted by machines. We all know from personal experience that the best managers and leaders take a holistic interest in their staff, taking the time and making the effort to understand the subtle, often intangible processes that connect their respective talents. As organizational management theorist Henry Mintzberg (2015) wisely puts it,

Measuring as a complement to managing is a fine idea: measure what you can; take seriously what you can't; and manage both thoughtfully. In other words: If you can't measure it, you'll have to manage it. If you can measure it, you'll especially have to manage it. Have we not had enough of leadership by remote control: sitting in executive offices and running the numbers—all that deeming and downsizing?²⁴

Contrary to the "what can't be measured can't be managed" idea, we *can* manage the less measurable if we embrace a wider set of tools and leave space for judgment. The key for practitioners is to begin with a recognition that measurability is not an indicator of significance and that professional management involves far more than simply "running the numbers," as Mintzberg puts it. Perhaps the most compelling empirical case for the importance of "navigating by judgment" in public administration has been made by Honig (2018), in which he shows—using a mix of quantitative data and case study analysis—that the more complex the policy intervention, the more necessary it becomes to grant discretionary space to frontline managers, and the more necessary such discretion is to achieving project success. Having ready access to relevant, high-quality quantitative data can aid in this "navigation," but true navigation requires access to a broader suite of empirical inputs.

In a similar vein, Ladner (2015, 3) points out that "standard performance monitoring tools are not suitable for highly flexible, entrepreneurial programs as they assume that how a program will be implemented follows its original design." To avoid "locking in" a theory of change that prevents exploration or responsive adaptation, some practitioners have provided helpful suggestions for how to use various planning frameworks in ways that support program learning.²⁵ The Building State Capability team highlights lighter-touch methods, such as their PDIA "check-ins," which include a series of probing questions

to assist teams in capturing learning and maximizing adaptation. Teskey and Tyrrel (2017) recommend participating in regularized formal and informal Review and Reflection (R&R) points, during which a contractor can demonstrate how politics, interests, incentives, and institutions were systematically considered in problem selection and design and, in turn, justify why certain choices were made to stop, drop, halt, or expand any activity or budget during implementation. The common connection across all these tools is that they seek to carve out meaningful space for qualitative data and the hard-won insights born out of practical experience.

In summary, then, public administrators can embed the recognition that management must be about more than measuring by first recognizing that whatever they choose to measure will inevitably create incentives to neglect processes and outcomes that cannot be measured (or are hard to measure) but are nonetheless crucial for discerning whether, how, where, and for whom policies are working. Following that recognition, they need to be very careful about what they choose to measure. Second, they can actively identify what they cannot (readily) measure that matters and take it seriously, developing strategies to manage that as well. A key part of those strategies will be that they create space for judgment, qualitative data inputs, and the practical experience of embedded individuals (focus group discussions, case studies, semi-structured interviews, review and reflection points, etc.) and treat these inputs as equally valid alongside more quantitative ones. As far as longer-term strategies to manage the immeasurable, administrations can work toward developing organizational systems that foster navigation. Such systems might include, for example, a management structure that delegates high levels of discretion to allow those on the ground the ability to navigate complex situations, recruitment strategies that foster high numbers of staff with extensive context-specific knowledge, and systems of monitoring and learning that encourage the routine evaluation of theory with practice.

CONCLUSION

Quantitative measurement in public administration is undoubtedly a critical arrow in the quiver of any attempt to improve the delivery of public services. And yet, since not everything that matters can be measured and not everything that can be measured matters, a managerial emphasis on measurement alone can quickly and inadvertently generate unwanted outcomes and unwarranted conclusions. In the everyday practices of public administration, effective and professional action requires forging greater complementarity between different epistemological approaches to collecting, curating, analyzing, and interpreting data. We fully recognize that this is easier said than done. The risks of reductive approaches to measurement are not unknown, and yet simplified appeals to “what gets measured gets managed” persist because they offer managers a form of escape from those “pesky human elements” that are difficult to understand and even more so to shift.

Most public administrators might agree in principle that a more balanced data suite is necessary to navigate their professional terrain, yet such aspirations are too often honored in the breach: under sufficient pressure to “deliver results,” staff from the top to the bottom of an organization are readily tempted to reverse engineer their behavior in accordance with what “the data” say (or can be made to say). Management as measurement is tempting for individuals and organizations that fear the vulnerability of their domain to unfavorable comparison with other (more readily measurable and “legible”) domains, as well as the complexity of problem solving and the necessity of subjective navigation that it often entails. But given how heavily institutional and sociopolitical factors shape how data are collected, how well they are collected and curated, and how they can be manipulated for unwarranted purposes, a simplistic approach to data as an easy fix is virtually guaranteed to obscure learning and hamper change efforts. If administrations genuinely wish to build their problem-solving capabilities, then access to more and better quantitative data will be necessary, but it will not be sufficient.

Beginning with an appreciation that much of what matters cannot be (formally) measured, public administration must routinely remind itself that promoting and accessing data is not an end in itself: data's primary purpose is not just monitoring processes, compliance, and outcomes, but contributing to problem solving and organizational learning. More and better data will not fix a problem if the absence of such data is not itself the key problem or the "binding constraint." Administrations that are committed to problem solving, therefore, will need to embed their measurement task in a broader problem-driven framework, integrate complementary qualitative data, and value embedded experience in order to apprehend and interpret complex realities more accurately. Their priority in undertaking good diagnostic work should be to identify and deconstruct key problems, using varied sources of data, and then to track and learn from potential solutions authorized and enacted in response to the diagnosis. Accurate inferences for policy and practice are not derived from data alone; close interaction is required between data (in various forms), theory, and experience. In doing all this, public administrators will help mitigate the distortionary (and ultimately self-defeating) effects of managing only that which is measured.

NOTES

Our thanks to Galileu Kim, Daniel Rogger, Christian Schuster, and participants at an authors' workshop for helpful comments and constructive suggestions. More than 20 years of collaboration with Vijayendra Rao have also deeply shaped the views expressed herein. Remaining errors of fact or interpretation are solely ours.

1. See, for example, former USAID Administrator Andrew Natsios (2011), citing Lord Wellington in 1812 on the insidious manner in which measures of "accountability" can compromise rather than enable central policy objectives (in Wellington's case, winning a war). For his part, Stiglitz has argued that "what you measure affects what you do. If you don't measure the right thing, you don't do the right thing" (quoted in Goodman 2009). Pritchett (2014), exemplifying this point, notes (at least at the time of his writing) that the Indian state of Tamil Nadu had 817 indicators for measuring the delivery of public education but none that actually assessed whether students were learning. In this instance, an abundance of "measurement" and "data" was entirely disconnected from (what should have been) the policy's central objective. In many cases, however, it is not always obvious, especially *ex ante*, what constitutes the "right thing" to measure—hence the need for alternative methodological entry points to elicit what this might be.
2. Social science methodology courses classically distinguish between four key issues that are at the heart of efforts to make empirical claims in applied research: *construct validity* (the extent to which any concept, such as "corruption" or "poverty," matches particular indicators), *internal validity* (the extent to which causal claims have controlled for potential confounding factors, such as sample selection bias), *external validity* (the likelihood that claims are generalizable at larger scales and to more diverse populations or novel contexts), and *reliability* (the extent to which similar findings would be reported if repeated or replicated by others). See, among many others, Johnson, Reynolds, and Mycoff (2019). Of these four issues, qualitative methods are especially helpful in ensuring construct validity, since certain terms may mean different things to different people in different places, complicating matters if one seeks to draw comparisons across different linguistic, cultural, or national contexts. In survey research, for example, it is increasingly common to include what is called an "anchoring vignette"—a short, real-world example of the phenomenon in question, such as an instance of corruption by a government official at a port—before asking the formal survey question so that cross-context variations in interpretation can be calibrated accordingly (see, among others, King and Wand 2007). Qualitative methods can also contribute to considerations pertaining to internal validity (Cartwright 2017) and external validity—helping to identify the conditions under which findings "there" might apply "here" (Woolcock 2018; see also Cartwright and Hardie 2012).
3. If such agencies or departments do in fact happen to perform especially strongly—in the spirit of the "positive deviance" cases of government performance in Ghana provided in McDonnell (2020)—then it would be useful to understand how and why this has been attained. For present purposes, our point is that, perhaps paradoxically, we should not expect, *ex ante*, that agencies or departments in the business of collecting and curating data for guiding policy and performance should themselves be exemplary exponents of the deployment of that data to guide their *own* performance—because doing this is a separate ontological task, requiring distinct professional capabilities. Like the proverbial doctors, if data analysts cannot "heal themselves," we should not expect other public agencies to be able to do so merely by "infusing them with more and better data."
4. A special issue of *The Journal of Development Studies*, 51.2, was dedicated to this problem. For example, on the enduring challenges associated with agricultural data—another sector with a long history of data collection experience—see Carletto, Jolliffe, and Banerjee (2015).

5. The adage is popularly known as GIGO: garbage in, garbage out.
6. See the evolution in early work on gender inclusion in rural India and subsequent work (Ban and Rao 2008; Duflo 2012; Sanyal and Rao 2018).
7. This does not mean, of course, that nothing can be said about GEPs after five years—managers and funders would surely want to know by this point whether the apparent “no net impact” claim is a result of poor technical design, weak implementation, contextual incompatibility, countervailing political pressures, or insufficient time having elapsed. Moreover, they would likely be interested in learning whether the GEP’s zero “average treatment effect” is nonetheless a process of offsetting outcomes manifest in a high standard deviation (meaning the GEP works wonderfully for some groups in some places but disastrously for others) and/or is yielding unanticipated or unmeasured outcomes (whether positive or negative). For present purposes, our point is that reliance on a single form and methodological source of data is unlikely to be able to answer these crucial administrative questions; with a diverse suite of methods and data, however, such questions become both askable and answerable. (See Rao, Ananthpur, and Malik 2017 for an instructive example, discussed below.)
8. One could say that this is a social scientific version of the Heisenberg uncertainty principle, in which the very act of measuring something changes it. See also Breckenridge (2014) on the politics and legacy of identity measurement in pre- and postcolonial South Africa and Hostetler (2021) on the broader manner in which imposing singular (but often alien) measures of time, space, and knowledge enabled colonial administration. More generally, Sheila Jasanoff’s voluminous scholarship shows how science is a powerful representation of reality, which, when harnessed to technology, can reduce “individuals to standard classifications that demarcate the normal from the deviant and authorize varieties of social control” (Jasanoff 2004, 13).
9. Among the classic historical texts on this issue are *Peasants into Frenchmen* (Weber 1976) and *Imagined Communities* (Anderson 1983). For more recent discussions, see Lewis (2015) on “the politics and consequences of performance measurement” and Beraldo and Milan (2019) on the politics of big data.
10. This is the finding, for example, from a major empirical assessment of cross-country differences regarding COVID-19 (Bollyky et al. 2022), wherein—controlling for a host of potential confounding variables—those countries with both high infections and high fatalities are characterized by low levels of trust between citizens and their governments and between each other. See further discussion of this study and its implications below.
11. The British movie *I, Daniel Blake* provides a compelling example of how even the literate in rich countries can be excluded by administrative systems and procedures that are completely alien to them—for example, filling out forms for unemployment benefits on the internet that require users to first “log on” and then “upload” a “CV.” The limits of formal measurement to bring about positive policy change has long been recognized; when the Victorian-era writer George Eliot was asked why she wrote novels about the lives of the downtrodden rather than contributing to official government reports more formally documenting their plight, she astutely explained that “appeals founded on generalizations and statistics require a sympathy ready-made, a moral sentiment already in activity” (quoted in Gill 1970, 10). Forging such Smithian “sympathy” and “moral sentiment” is part of the important antecedent work that renders “generalizations and statistics” legible and credible to those who might otherwise have no reason for engaging with, or experience interpreting, such encapsulations of reality.
12. We fully recognize that, in principle, econometricians have methods available to identify both outcome heterogeneity and the factors driving it. Even so, if local average treatment effects are reported as zero, the “no impact” conclusion is highly likely to be the (only) key takeaway message. The primary benefit of incorporating both qualitative and econometric methods is the capacity of the former to identify factors that were not anticipated in the original design (see Rao 2022). In either case, Ravallion’s (2001) injunction to “look beyond averages” when engaging with complex phenomena is worth being heeded by all researchers (and those that interpret researchers’ findings), no matter their disciplinary or methodological orientations.
13. On the use of mixed methods in process evaluations, see Rogers and Woolcock (2023).
14. The six pillars were: maternal, newborn, and child health; childhood essential medicines and increasing treatment of important childhood diseases; improving child nutrition; immunization; malaria control; and the elimination of mother-to-child transmission of human immunodeficiency virus (HIV).
15. A PforR is one of the World Bank’s three financing instruments. Its unique features are that it uses a country’s own institutions and processes and links disbursement of funds directly to the achievement of specific program results. Where “traditional” development interventions proceed on the basis of ex ante commitments (for example, to designated “policy reforms” or to the adoption of procedures compliant with international standards), PforR-type interventions instead reward the attainment of predetermined targets, typically set by extrapolating from what recent historical trajectories have attained. According to the Project Appraisal Document for SOML, “each state would be eligible for a grant worth \$325,000 per the percentage point gain they made above average annual gain in the sum of six indicators of health service coverage.” The six indicators were: vitamin A, Pentavalent3 immunization, use of insecticide-treated nets (ITNs) by children under five, skilled birth attendance, contraceptive prevalence rate, and the prevention of mother-to-child transmission of HIV.
16. These tables are based on student performance in standardized tests in mathematics, reading, and science, which are administered by the Paris-based Organisation for Economic Co-operation and Development (OECD).

17. A summary of the report explains that

whereas negotiation and flexibility are high on the list of skills for 2015, in 2020 they will begin to drop from the top 10 as machines, using masses of data, begin to make our decisions for us. A survey done by the World Economic Forum's Global Agenda Council on the Future of Software and Society shows people expect artificial intelligence machines to be part of a company's board of directors by 2026. Similarly, active listening, considered a core skill today, will disappear completely from the top 10. Emotional intelligence, which doesn't feature in the top 10 today, will become one of the top skills needed by all. (Gray 2016)

See also Soffel (2016).

18. Many companies and tertiary institutions are ahead of the curve in this regard. Recently, over 150 of the top private high schools in the US, including Phillips Exeter Academy and the Dalton School—storied institutions that have long relied on the status conveyed by student ranking—have pledged to shift to new transcripts that provide more comprehensive, qualitative feedback on students while ruling out any mention of credit hours, GPAs, or A–F grades. And colleges—the final arbiters of high school performance—are signaling a surprising willingness to depart from traditional assessments that have been in place since the early 19th century. From Harvard and Dartmouth to small community colleges, more than 70 US institutions of higher learning have weighed in, signing formal statements asserting that competency-based transcripts will not hurt students in the admissions process. See the “College Admissions” page on the New England Secondary School Consortium website: <http://www.newenglandssc.org/resources/college-admissions/>.
19. See Milante and Woolcock (2017) for a complementary set of dynamic quantitative and qualitative measures by which a given country might be declared a “fragile” state.
20. For development-oriented organizations, a set of tools and guidelines for guiding this initial assessment according to a political economy analysis (PEA) framework—crafted by USAID and ODI (London) and adopted by certain parts of the World Bank—is *Thinking and Working Politically through Applied Political Economy Analysis: A Guide for Practitioners* (Rocha Menocal et al. 2018). Its key observations include the following. First, a well-designed process of policy implementation should answer not only the technical question of what needs to be done but also how it should be done. Second, in-depth understanding of the political, economic, social, and cultural forces needs to supplement technical analysis to achieve successful policy implementation. Third, PEA should incorporate three pillars: the foundational factors (geography, natural resource occurrence, national borders), the “rules of the game” (institutions at the formal [political system, administrative structure, and law] and the informal [social and cultural norms] levels), and the “here and now” (current leaders, geopolitical situation, and natural hazards). Fourth, it is crucial to pay attention to the institutions, the structure of incentives, and the constraints, as well as the gains and losses of all the actors involved in policy implementation, including those outside of the traditional purview of development organizations. Fifth, policy solutions should be adjusted to political realities encountered on the ground in an iterative and incremental fashion. And finally, the evaluation of policy success should be extended to incorporate “process-based indicators,” including trust and quality of relationship. Hudson, Marquette, and Waldock (2016) offer a guide for “everyday political analysis,” which introduces a stripped-back political-analysis framework designed to help frontline practitioners make quick but politically informed decisions. It aims to complement more in-depth political analysis by helping programming staff to develop the “craft” of political thinking in a way that fits their everyday working practices.
21. On the application of such efforts to the case of policing in particular, see Sparrow (2018).
22. The *PDI* toolkit: *A DIY Approach to Solving Complex Problems* (Samji et al. 2018) was designed by members of Harvard's Building State Capability program to guide government teams through the process of identifying, deconstructing, and solving complex problems. See in particular the section “Constructing your problem,” which guides practitioners through the process of defining a problem that matters and building a credible, measurable vision of what success would look like.
23. As anthropologist Mike McGovern (2011, 353) powerfully argues, taking context seriously

is neither a luxury nor the result of a kind of methodological altruism to be extended by the soft-hearted. It is, in purely positivist terms, the epistemological due diligence work required before one can talk meaningfully about other people's intentions, motivations, or desires. The risk in foregoing it is not simply that one might miss some of the local color of individual “cases.” It is one of misrecognition. Analysis based on such misrecognition may mistake symptoms for causes, or two formally similar situations as being comparable despite their different etiologies. To extend the medical metaphor one step further, misdiagnosis is unfortunate, but a flawed prescription based on such a misrecognition can be deadly.

More generally, see Hoag and Hull (2017) for a summary of the anthropological literature on the civil service. Bailey (2017) provides a compelling example of how insights from qualitative fieldwork help explain the strong preference among civil servants in Tanzania for providing new water infrastructure projects over maintaining existing ones. Though a basic benefit-cost analysis favored prioritizing maintenance, collective action problems among civil servants themselves, the prosaic challenges of mediating local water management disputes overseen by customary institutions, and the performance targets set by the government all conspired to create suboptimal outcomes.

24. Sainsbury (2015): “Someone I know once asked a most senior British civil servant why his department had to do so much measuring. His reply: ‘What else can we do when we don’t know what’s going on?’ Did he ever try getting on the ground to find out what’s going on? And then using judgment to assess that?”
25. Teskey (2017) and Wild, Booth, and Valters (2017) give examples of an adaptive logframe, drawn from Department for International Development experiences, that sets out clear objectives at the outcome level and focuses monitoring of outputs on the quality of the agreed rapid-cycle learning process. Strategy Testing (ST) is a monitoring system that the Asia Foundation developed specifically to track programs that are addressing complex development problems through a highly iterative, adaptive approach.

REFERENCES

- Anderson, Benedict. 1983. *Imagined Communities: Reflections on the Origins and Spread of Nationalism*. London: Verso.
- Andrews, Matt. 2013. *The Limits of Institutional Reform in Development: Changing Rules for Realistic Solutions*. New York: Cambridge University Press.
- Andrews, Matt, Lant Pritchett, and Michael Woolcock. 2017. *Building State Capability: Evidence, Analysis, Action*. New York: Oxford University Press.
- Bailey, Julia. 2017. “Bureaucratic Blockages: Water, Civil Servants and Community in Tanzania.” Policy Research Working Paper 8101, World Bank, Washington, DC.
- Ban, Radu, and Vijayendra Rao. 2008. “Tokenism or Agency? The Impact of Women’s Reservations on Village Democracies in South India.” *Economic Development and Cultural Change* 56 (3): 501–30. <https://doi.org/10.1086/533551>.
- Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukherji, Marc Shotland, and Michael Walton. 2016. “Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of ‘Teaching at the Right Level’ in India.” NBER Working Paper 22746, National Bureau of Economic Research, Cambridge, MA.
- Beraldo, Davide, and Stefania Milan. 2019. “From Data Politics to the Contentious Politics of Data.” *Big Data & Society* 6 (2): 1–11. <https://doi.org/10.1177/2053951719885967>.
- Bollyky, Thomas J., Erin N. Hulland, Ryan M. Barber, James K. Collins, Samantha Kiernan, Mark Moses, David M. Pigott, et al. 2022. “Pandemic Preparedness and Covid-19: An Exploratory Analysis of Infection and Fatality Rates, and Contextual Factors Associated with Preparedness in 177 Countries, from Jan 1, 2020, to Sept 30, 2021.” *The Lancet* 399 (10334): 1489–512. [https://doi.org/10.1016/S0140-6736\(22\)00172-6](https://doi.org/10.1016/S0140-6736(22)00172-6).
- Breckenridge, Keith. 2014. *Biometric State: The Global Politics of Identification and Surveillance in South Africa, 1850 to the Present*. Cambridge, UK: Cambridge University Press.
- Bridges, Kate, and Michael Woolcock. 2017. “How (Not) to Fix Problems That Matter: Assessing and Responding to Malawi’s History of Institutional Reform.” Policy Research Working Paper 8289, World Bank, Washington, DC.
- Bridges, Kate, and Michael Woolcock. 2019. “Implementing Adaptive Approaches in Real World Scenarios: A Nigeria Case Study, with Lessons for Theory and Practice.” Policy Research Working Paper 8904, World Bank, Washington, DC.
- Cahill, Jonathan. 2017. *Making a Difference in Marketing: The Foundation of Competitive Advantage*. London: Routledge.
- Carletto, Calogero, Dean Jolliffe, and Raka Banerjee. 2015. “From Tragedy to Renaissance: Improving Agricultural Data for Better Policies.” *The Journal of Development Studies* 51 (2): 133–48. <https://doi.org/10.1080/00220388.2014.968140>.
- Cartwright, Nancy. 2017. “Single Case Causes: What Is Evidence and Why.” In *Philosophy of Science in Practice*, edited by Hsiang-Ke Chao and Julian Reiss, 11–24. New York: Springer.
- Cartwright, Nancy, and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. New York: Oxford University Press.
- Caulkin, Simon. 2008. “The Rule Is Simple: Be Careful What You Measure.” *Guardian* (US edition), February 9, 2008. <https://www.theguardian.com/business/2008/feb/10/businesscomment1>.
- Caulkin, Simon. 2016. “Decision-Making: How to Make the Most of Data.” *The Treasurer*, January 4, 2016. <https://www.treasurers.org/hub/treasurer-magazine/decision-making-how-make-most-data>.
- Claxton, Guy, and Bill Lucas. 2015. *Educating Ruby: What Our Children Really Need to Learn*. New York: Crown House Publishing.
- Dirks, Nicholas. 2011. *Castes of Mind*. Princeton, NJ: Princeton University Press.
- Dom, Catherine, Alasdair Fraser, Joseph Holden, and John Patch. 2020. “Results-Based Financing in the Education Sector: Country-Level Analysis. Final Synthesis Report.” Report submitted to the REACH Program at the World Bank by Mokoro Ltd.

- Duflo, Esther. 2012. "Women Empowerment and Economic Development." *Journal of Economic Literature* 50 (4): 1051–79. <https://doi.org/10.1257/jel.50.4.1051>.
- Economist*. 2021. "How World Bank Leaders Put Pressure on Staff to Alter a Global Index." September 17, 2021. <https://www.economist.com/finance-and-economics/2021/09/17/how-world-bank-leaders-put-pressure-on-staff-to-alter-a-global-index>.
- Gill, Stephen. 1970. "Introduction to Elizabeth Gaskell." In *Mary Barton: A Tale of Manchester Life*. London: Penguin.
- Global Partnership for Education. 2020. *21st-Century Skills: What Potential Role for the Global Partnership for Education?* Washington, DC: Global Partnership for Education Secretariat. <https://www.globalpartnership.org/sites/default/files/document/file/2020-01-GPE-21-century-skills-report.pdf>.
- Goodman, Peter S. 2009. "Emphasis on Growth Is Called Misguided." *New York Times*, October 4, 2009. <https://www.nytimes.com/2009/09/23/business/economy/23gdp.html>.
- Gray, Alex. 2016. "The 10 Skills You Need to Thrive in the Fourth Industrial Revolution." World Economic Forum. <https://www.weforum.org/agenda/2016/01/the-10-skills-you-need-to-thrive-in-the-fourth-industrial-revolution>.
- 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.
- Honig, Dan. 2018. *Navigation by Judgment: Why and When Top-Down Management of Foreign Aid Doesn't Work*. New York: Oxford University Press.
- Honig, Dan, and Lant Pritchett. 2019. "The Limits of Accounting-based Accountability in Education (and Far Beyond): Why More Accounting Will Rarely Solve Accountability Problems." Working Paper No. 510, Center for Global Development, Washington, DC.
- Hostetler, Laura. 2021. "Mapping, Registering, and Ordering: Time, Space and Knowledge." In *The Oxford World History of Empire: Volume One: The Imperial Experience*, edited by Peter Fibiger Bang, C. A. Bayly, and Walter Scheidel, 288–317. New York: Oxford University Press.
- Hudson, David, Heather Marquette, and Sam Waldo. 2016. "Everyday Political Analysis." Working paper, Developmental Leadership Program, University of Birmingham, Birmingham, UK.
- Jasanoff, Sheila. 2004. "Ordering Knowledge, Ordering Society." In *States of Knowledge: The Co-Production of Science and the Social Order*, edited by Sheila Jasanoff, 13–45. London: Routledge.
- Jerven, Morten. 2013. *Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It*. Ithaca, NY: Cornell University Press.
- Johnson, Janet Buttolph, Henry T. Reynolds, and Jason D. Mycoff. 2019. *Political Science Research Methods*. 9th ed. Thousand Oaks, CA: Sage.
- Khan, Salman. 2012. *The One World Schoolhouse: Education Reimagined*. London: Hodder & Stoughton.
- King, Gary, and Jonathan Wand. 2007. "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes." *Political Analysis* 15 (1): 46–66. <https://doi.org/10.1093/pan/mpl011>.
- Ladner, Debra. 2015. *Strategy Testing: An Innovative Approach to Monitoring Highly Flexible Aid Programs*. Working Politically in Practice Case Study 3. San Francisco: The Asia Foundation.
- Lewis, Jenny M. 2015. "The Politics and Consequences of Performance Measurement." *Policy and Society* 34 (1): 1–12. <https://doi.org/10.1016/j.polsoc.2015.03.001>.
- Mamdani, Mahmood. 2002. *When Victims Become Killers: Colonialism, Nativism, and the Genocide in Rwanda*. Princeton, NJ: Princeton University Press.
- McCarthy, Niall. 2019. "The Countries Best Prepared to Deal With a Pandemic." *Statista*, October 28, 2019. <https://www.statista.com/chart/19790/index-scores-by-level-of-preparation-to-respond-to-an-epidemic/>.
- McDonnell, Erin. 2017. "Patchwork Leviathan: How Pockets of Bureaucratic Governance Flourish within Institutionally Diverse Developing States." *American Sociological Review* 82 (3): 476–510. <https://doi.org/10.1177/0003122417705874>.
- McDonnell, Erin. 2020. *Patchwork Leviathan: Pockets of Bureaucratic Effectiveness in Developing States*. Princeton, NJ: Princeton University Press.
- McGovern, Mike. 2011. "Popular Development Economics: An Anthropologist among the Mandarins." *Perspectives on Politics* 9 (2): 345–55. <https://doi.org/10.1017/S1537592711000594>.
- Merry, Sally Engle, Kevin E. Davis, and Benedict Kingsbury, eds. 2015. *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law*. New York: Cambridge University Press.
- Milante, Gary, and Michael Woolcock. 2017. "New Approaches to Identifying State Fragility." *Journal of Globalization and Development* 8 (1): 20170008. <https://doi.org/10.1515/jgd-2017-0008>.
- Mintzberg, Henry. 2015. "If You Can't Measure It, You'd Better Manage It." *Henry Mintzberg* (blog). May 28, 2015. <https://mintzberg.org/blog/measure-it-manage-it>.
- Natsios, Andrew. 2011. "The Clash of the Counter-Bureaucracy and Development." Essay, Center for Global Development, Washington, DC. https://www.cgdev.org/sites/default/files/1424271_file_Natsios_Counterbureaucracy.pdf.

- Porter, Theodore M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Pritchett, Lant. 2013. *The Rebirth of Education: Schooling Ain't Learning*. Washington, DC: Center for Global Development.
- Pritchett, Lant. 2014. "The Risks to Education Systems from Design Mismatch and Global Isomorphism: Concepts, with Examples from India." WIDER Working Paper 2014/039, United Nations University World Institute for Development Economics Research, Helsinki.
- Propper, Carol, Simon Burgess, and Denise Gossage. 2008. "Competition and Quality: Evidence from the NHS Internal Market 1991–9." *The Economic Journal* 118 (525): 138–70. <https://doi.org/10.1111/j.1468-0297.2007.02107.x>.
- Rao, Vijayendra. 2022. "Can Economics Become More Reflexive? Exploring the Potential of Mixed-Methods." Policy Research Working Paper 9918, World Bank, Washington, DC.
- Rao, Vijayendra, Kripa Ananthpur, and Kabir Malik. 2017. "The Anatomy of Failure: An Ethnography of a Randomized Trial to Deepen Democracy in Rural India." *World Development* 99 (11): 481–97. <https://doi.org/10.1016/j.worlddev.2017.05.037>.
- Ravallion, Martin. 2001. "Growth, Inequality and Poverty: Looking Beyond Averages." *World Development* 29 (11): 1803–15. [https://doi.org/10.1016/S0305-750X\(01\)00072-9](https://doi.org/10.1016/S0305-750X(01)00072-9).
- Ravallion, Martin. 2020. "Pandemic Policies in Poor Places." CGD Note, April 24, Center for Global Development, Washington, DC.
- Ridgway, V. F. 1956. "Dysfunctional Consequences of Performance Measurements." *Administrative Science Quarterly* 1 (2): 240–7. <https://doi.org/10.2307/2390989>.
- Robinson, Ken, and Lou Aronica. 2015. *Creative Schools: Revolutionizing Education from the Ground Up*. London: Penguin UK.
- Rocha Menocal, Alina, Marc Cassidy, Sarah Swift, David Jacobstein, Corinne Rothblum, and Ilona Tservil. 2018. *Thinking and Working Politically through Applied Political Economy Analysis: A Guide for Practitioners*. Washington, DC: USAID, DCHA Bureau Center of Excellence on Democracy, Human Rights, and Governance. https://usaidealarninglab.org/sites/default/files/resource/files/pea_guide_final.pdf.
- Rogers, Patricia, and Michael Woolcock. 2023. "Process and Implementation Evaluation Methods." In *Oxford Handbook of Program Design and Implementation*, edited by Anu Rangarajan, 294–316. New York: Oxford University Press.
- Samji, Salimah, Matt Andrews, Lant Pritchett, and Michael Woolcock. 2018. *PDIAtoolkit: A DIY Approach to Solving Complex Problems*. Cambridge, MA: Building State Capability, Center for International Development, Harvard University. <https://bsc.cid.harvard.edu/PDIAtoolkit>.
- Sandefur, Justin, and Amanda Glassman. 2015. "The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics." *The Journal of Development Studies* 51 (2): 116–32. <https://doi.org/10.1080/00220388.2014.968138>.
- Sanyal, Paromita, and Vijayendra Rao. 2018. *Oral Democracy: Deliberation in Indian Village Assemblies*. New York: Cambridge University Press.
- Scott, James C. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, CT: Yale University Press.
- Soffel, Jenny. 2016. "The 21st-Century Skills Every Student Needs." World Economic Forum. <https://www.weforum.org/agenda/2016/03/21st-century-skills-future-jobs-students/>.
- Sparrow, Malcolm. 2018. "Problem-Oriented Policing: Matching the Science to the Art." *Crime Science* 7 (1): 1–10. <https://doi.org/10.1186/s40163-018-0088-2>.
- Teskey, Graham. 2017. "Thinking and Working Politically: Are We Seeing the Emergence of a Second Orthodoxy?" Governance Working Paper Series 1, ABT Associates, Canberra.
- Teskey, Graham, and Lavinia Tyrrel. 2017. "Thinking and Working Politically in Large, Multi-Sector Facilities: Lessons to Date." Governance Working Paper Series 2, ABT Associates, Canberra.
- Weber, Eugen. 1976. *Peasants into Frenchmen: The Modernization of Rural France, 1870–1914*. Palo Alto, CA: Stanford University Press.
- Wild, Leni, David Booth, and Craig Valters. 2017. *Putting Theory into Practice: How DFID Is Doing Development Differently*. London: ODI.
- WilmerHale. 2021. *Investigation of Data Irregularities in "Doing Business 2018" and "Doing Business 2020": Investigation Findings and Report to the Board of Executive Directors*. <https://thedocs.worldbank.org/en/doc/84a922cc9273b7b120d49ad3b9e9d3f9-0090012021/original/DB-Investigation-Findings-and-Report-to-the-Board-of-Executive-Directors-September-15-2021.pdf>.
- Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.

- Woolcock, Michael. 2018. "Reasons for Using Mixed Methods in the Evaluation of Complex Projects." In *Contemporary Philosophy and Social Science: An Interdisciplinary Dialogue*, edited by Michiru Nagatsu and Attilia Ruzzene, 149–71. London: Bloomsbury Academic.
- World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank.
- World Economic Forum. 2016. *New Vision for Education: Fostering Social and Emotional Learning through Technology*. Industry Agenda. Geneva: World Economic Forum.
- Zak, Paul. 2013. "Measurement Myopia." Drucker Institute, Claremont, CA. December 6, 2021. <https://www.drucker.institute/thedx/measurement-myopia/>.
- Zhao, Yong. 2012. "Test Scores vs. Entrepreneurship: PISA, TIMSS, and Confidence." *Yong Zhao* (blog). <http://zhaolearning.com/2012/06/06/test-scores-vs-entrepreneurship-pisa-timss-and-confidence/>.

CHAPTER 5

Practical Tools for Effective Measurement and Analytics

Maria Ruth Jones and Benjamin Daniels

SUMMARY

Increasingly important features of good data analysis are the transparency with which the analysis is undertaken and the reproducibility of its results. These features ensure the credibility of analytical outputs and policy guidance. The World Bank's Development Impact Evaluation (DIME) Department has developed freely available tools and processes to support the achievement of these best practices by analysts across the world. These resources include research-cycle frameworks, extensive training tools, detailed archives of process and technical guidance, and a collaborative approach to data and analytics. The DIME Analytics team continuously updates many of these resources and makes them available globally as a free knowledge product. This chapter describes the frameworks, the approach, and the products that are available to bring these best practices into any organization that relies on data analytics for decision-making. The chapter provides a discussion of how to apply these elements to public administration, thus ensuring that analytics of government accord with international best practices.

ANALYTICS IN PRACTICE

- For credibility, modern methods for data analysis rely on analysts undertaking their work in a transparent way that ensures their results can be replicated.
- Best practices for reproducibility and transparency assure the internal quality and organization of data work and provide a template for publishing materials externally when appropriate.
- Producing analysis that accords with best practices requires considering the full life cycle of data work, such that each stage of handling data can be designed to support the next stages.

Maria Ruth Jones is a senior economist for the World Bank's Development Impact Evaluation (DIME) Department, where she coordinates DIME Analytics. Benjamin Daniels is a research fellow in the *gui*²de group at the McCourt School of Public Policy at Georgetown University, who works with DIME Analytics.

- Development Impact Evaluation (DIME) Analytics, a unit of the DIME Department at the World Bank, has developed extensive, publicly available tools and trainings to standardize and promulgate these best practices among analysts across the world.
- This chapter describes these resources and links to their locations online. It provides a discussion of how to apply these elements to public administration, thus ensuring government analytics accord with international best practices.

INTRODUCTION

The principles of good data analysis and empirical research have been refined over centuries. Today's analysts and researchers work in an environment in which adherence to modern standards of analysis is an essential part of the credibility of their results. As such standards become more abundant and layered, it can be a challenge to keep up to date on best practices. These considerations go beyond the internal analysis of government data and are of general concern to anyone attempting to undertake rigorous analysis.

To support analysts and researchers across the world in understanding and implementing modern approaches to data analysis, the World Bank's Development Impact Evaluation (DIME) Department has created a series of resources to support the adoption of best practices in data collection and analysis. The DIME Analytics unit aims to help data analysts identify inefficiencies and practices that compromise the quality of analysis and to develop a workflow that strengthens their work. DIME Analytics creates and tests tools that support these practices and provides the training and technical support necessary to sustain their adoption (illustrated in figure 5.1). This chapter aims to provide an introduction to the resources (typically free) DIME Analytics provides, which are in themselves an introduction to the global frontier of best practices in data analytics.

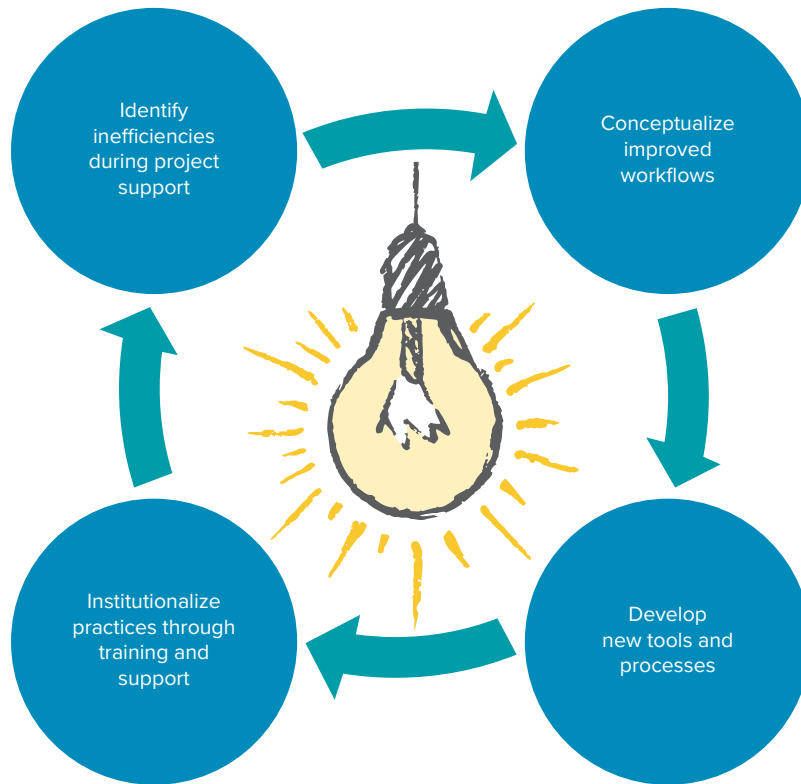
Creating high-quality data and research outputs is central to generating useful insights to inform public administration. DIME Analytics' resources are well suited to those interested in updating their administration's capacity to collect high-quality data and evaluate the impact of public administration reforms. This chapter therefore describes protocols for high-quality data collection relevant to all public administration data collection efforts and discusses how best practices for reproducible and transparent data collection and analysis can establish the credibility of public administration analysis and research.

ESTABLISHING THE CREDIBILITY OF PUBLIC ADMINISTRATION MEASUREMENT

The framework of reproducibility, transparency, and credibility has become the basis for empirical research and data analysis worldwide. There are several core reasons for the adoption of these principles in the academic world, and most of them apply equally to the sphere of policy analysis, design, and governance. In academic research, collaborators are often not employed in the same organization, or direct collaboration may never occur. These standards have been adopted so that disparate teams and individuals can access, reuse, assess, and build upon the work done by other researchers in the ecosystem, even when they do not have institutional or personal connections to one another. A similar argument can be made of the many distinct units of government administrations.

Reproducible data work is designed so that the materials and processes from a task can be easily reused and adapted within and across contexts. Such data work is done with the goals of collaboration and the handoff of materials in mind, even if the research materials are not intended for public consumption. When analytical processes are designed for collaboration and reuse, they become a form of knowledge

FIGURE 5.1 DIME Analytics Uses an Iterative Process to Expand Technical Capacity throughout the Research Cycle



Source: DIME (Development Impact Evaluation) Analytics, World Bank.

accumulation within an organization. By contrast, data processes that are completed as “one-off” tasks—nonreproducible processes that cannot be transferred to other people after completion—cannot transfer workflow improvements and efficiencies to other tasks. Reproducible data work also enables more people to work on a task because the tools can be handed off among individuals on the team; it also enables more people to conduct quality control over the work of others, which is a built-in feature of reusing others’ work.

Transparent work is designed so the materials and processes from a task can be understood and used by individuals not involved in the original task. In research, this often means publishing or releasing code, data, and documentation to the general public (or under license to other researchers). However, publication is not the most important aspect of transparent work. The essential characteristic of transparent work is that the documentation of all steps is complete, the organization of subtasks is easily understood, and the complete materials are archived in a known location. In this way, a transparent process is one that can be accessed and replicated by a new team without contacting the original analysts. This approach allows knowledge transfer to happen rapidly both across an organization, from team to team, and within an organization as materials are reused and improved upon progressively over time.

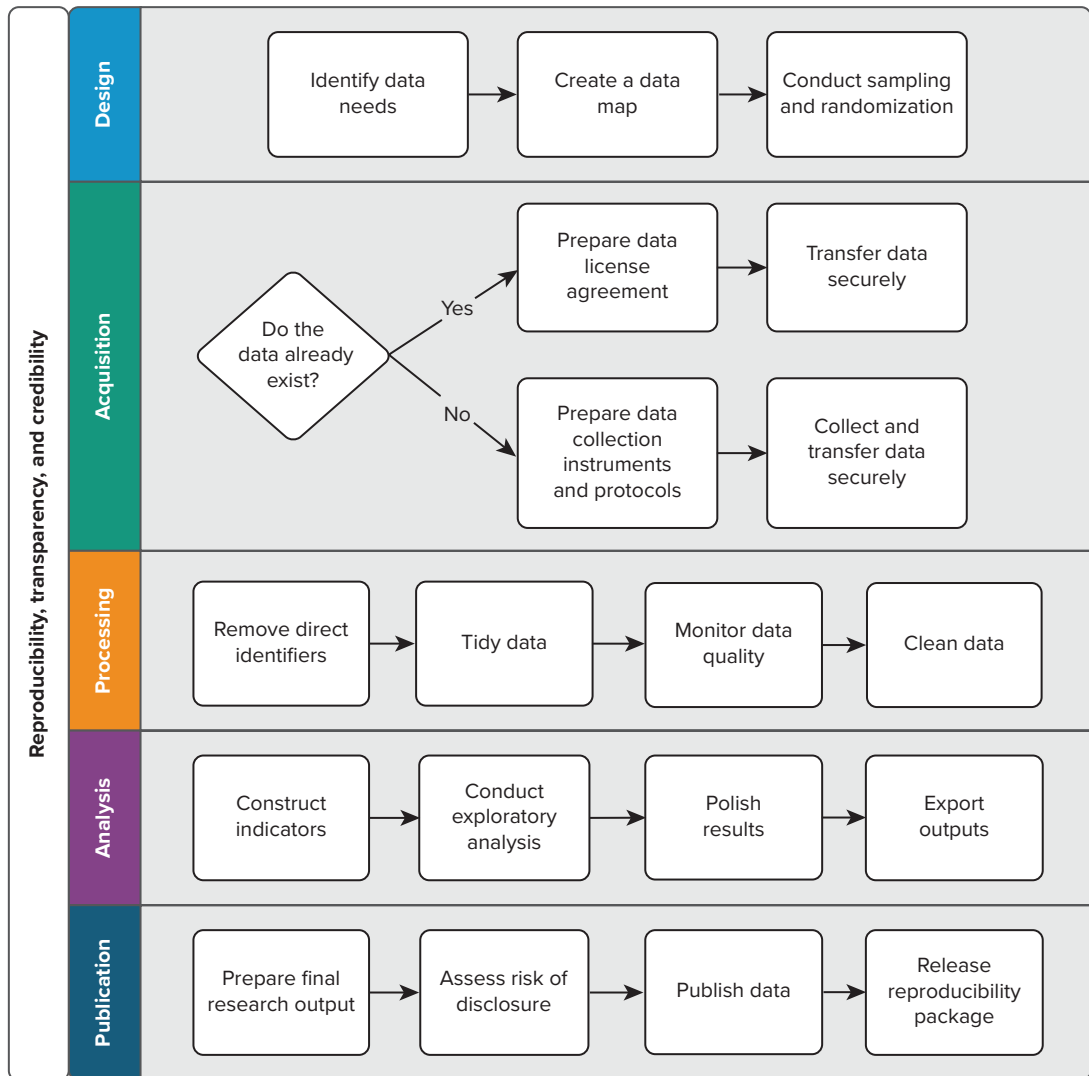
Together, these two approaches produce a foundation of credibility for conclusions and outputs within an organization. Reproducible work usually demands the adoption of standardized processes for subtasks. Processes that are reused frequently and by many people will attract higher levels of investment and quality assurance. Similarly, when data, code, and documentation are intended to be seen and reused by others, authors are incentivized to improve their own organization and description of materials, particularly if they or their own team will be the ones reusing them in the future. When these materials are archived and are visible to other members

of an organization, they can be quality-controlled, and, over time, the best work becomes widely improved and adopted, establishing the reputation of high-quality tools and processes. Whether or not analytical materials are ever released to the public is secondary to adopting a mindset of collaboration and reuse across a research organization.

PROTOCOLS FOR HIGH-QUALITY DATA COLLECTION

The conventions, standards, and best practices that are fast becoming a necessity for high-quality empirical research affect most elements of data collection and analysis. Figure 5.2 describes the standard workflow of analysis from design to reporting or publication. Some of these steps may be redundant, such as when using administrative data, but considering the full workflow allows analysts to imagine what the optimal process would have been if they had been in control of each stage of data collection and analysis.

FIGURE 5.2 Overview of the Tasks Involved in Development Research Data Work



Source: Bjärkefur et al. 2021.

DIME Analytics has described best practices at each of these stages in its published handbook, *Development Research in Practice: The DIME Analytics Data Handbook* (Bjærkefur et al. 2021). *The DIME Analytics Data Handbook* is intended to train users of data in effective, efficient, and ethical data handling. It covers the full data workflow for an empirical project, compiling all the lessons and tools developed by DIME Analytics into a single narrative of best practices. It provides a step-by-step guide to high-quality, reproducible data work at each stage of the data workflow, from design to publication, as visualized in figure 5.2. Each chapter contains boxes with examples of how the practices and workflows described in that chapter were applied in a real-life case. The handbook is available for free download through the World Bank’s Open Knowledge Repository or for purchase from Amazon. The handbook and related resources will be updated over time as best practices evolve, and feedback can always be provided through the handbook’s GitHub repository.¹

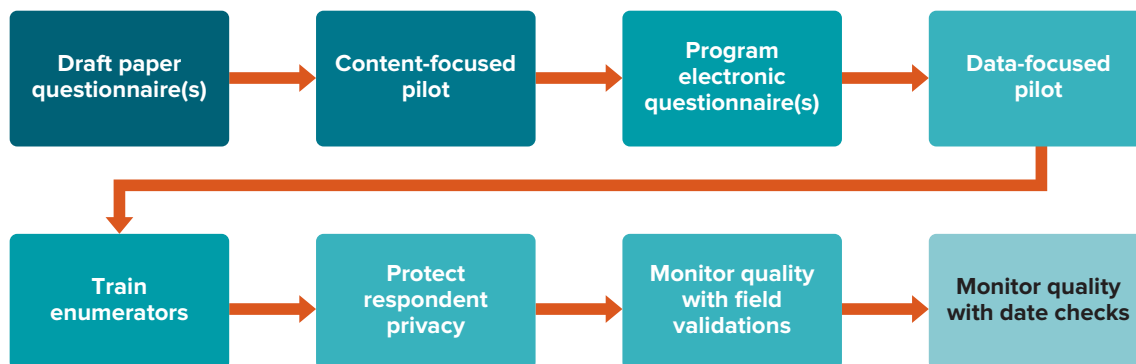
Let us take as an example the element “Prepare data collection instruments and protocols” in the “Acquisition” pillar in figure 5.2. This activity can be further broken down into constituent parts. The handbook provides narrative descriptions and best practices for each stage of the activity. It also provides links to corresponding entries on the DIME Wiki, which contain more specific technical details and are kept up to date as these details change and evolve.² Figure 5.3 illustrates the full workflow for an electronic survey.

Each of these stages can then be broken down further into constituent activities. Figure 5.4 summarizes the key protocols for ensuring data quality at every stage of the survey workflow. In this way, for each activity related to data collection, DIME Analytics has outlined the activities required to undertake rigorous data collection and analysis and has provided a knowledge network for analysts to obtain the appropriate level of specific detail about each task and subtask for their current needs. Adhering to such standardized and organized protocols, even when analysis is not intended to be made public, ensures that work is organized and internally consistent. In this way, the DIME Analytics knowledge products reflect the world’s frontier knowledge about how to obtain credible empirical results.

As an example of the application of these principles in public administration, consider an assessment of the impact of in-service training. Much effort is made to keep the skills of public servants current with new procedures within government and innovations in wider society. However, there is frequently little evaluation of whether such training actually has an impact on the quality of government processes or productivity. An effective measurement framework might include an immediate assessment of public servants’ skills upon entry to and exit from the training, as well as a follow-up with objective assessments of the quality of procedure and productivity.

To assess the broad impact of an in-service training intervention, we could imagine returning to the unit of the trainee after some time had passed and assessing the impact of the training through a survey

FIGURE 5.3 Workflow for High-Quality Survey Data Collection



Source: DIME (Development Impact Evaluation) Analytics, World Bank.

FIGURE 5.4 Summary Protocols for High-Quality Data Collection at Every Stage of the Survey Workflow



Source: DIME (Development Impact Evaluation) Analytics, World Bank.

of the unit's public servants. Working through the elements of figure 5.4, this would require the following steps:

- **Prepare for a survey:** Ensure that we have a sufficiently large sample of trainees to detect a quantitative increase in their knowledge. Determine which organizations or units we want to survey and any sub-groups of respondents we want to focus on (managers vs. nonmanagers, contract workers vs. staff, etc.). Obtain organizational structure charts if necessary.
- **Draft survey instrument:** Develop questions sufficiently close to the concepts of interest to provide us with measures we can act on. We might ask the trainees questions about how they have used the new

procedures in their daily work, and we might ask their colleagues whether this has made them easier to work with on related tasks. Basing our survey on existing surveys of these trainees would allow us to track the impacts of training across time.

- **Pilot:** Test intended questions with the audience of interest so the trainees understand what we are asking about. Though we might be asking about the actual practice of using the new procedures, the wrong question will make trainees respond about the change in rules rather than whether they are actually being followed.
- **Train enumerators:** If data collection is managed by enumerators, ensure they understand the appropriate protocols of the public service and take a consistent approach to measurement. If trainees differ in seniority, how will an enumerator ensure a common approach to surveying across the hierarchy?
- **Protect respondents' privacy:** Create protocols to ensure the process protects respondents' privacy. Individual public servants' capability at aspects of their jobs is highly sensitive information. If officials believe that their data will not be completely private, they may refuse to cooperate with the data collection exercise.
- **Monitor quality with field validations and/or data checks:** Implement a system that monitors survey data as they come in, and monitor whether a specific enumerator, department, or agency is presenting unusual data.

For a more detailed explanation of the protocols associated with surveys, please refer to chapter 4 of *The DIME Analytics Data Handbook* (Björkefur et al. 2021) and the videos and lectures from the Manage Successful Impact Evaluation Surveys course (DIME Analytics 2020a, 2021b).³ To learn how to implement high-quality surveys in practice, please refer to the DIME Wiki articles on Primary Data Collection, Field Surveys, Survey Protocols, and Remote Surveys.⁴ For specific considerations regarding phone surveys, please refer to chapters 1 and 2 of *Mobile Phone Panel Surveys in Developing Countries: A Practical Guide for Microdata Collection* (Dabalen et al. 2016).

PUBLIC RESOURCES AND TOOLS

To support the implementation of a rigorous research process, DIME Analytics has made a range of resources, technical solutions, and research protocols available through open-access trainings and open-source tools. We have found that there is significant unmet demand for these public goods, demonstrated by fast-growing and widespread global interest in our offerings. This section describes DIME Analytics' flagship resources.

Development Research in Practice: The Course

The DIME Analytics Data Handbook is accompanied by the course Development Research in Practice.⁵ This free and fully virtual course lasts 8 weeks, with seven lecture weeks each corresponding to one of the chapters from the handbook. This course provides attendees with a high-level overview of the entire process of empirical research so that they understand how each stage of the research workflow fits among the others and how the framework of transparency, reproducibility, and credibility informs the entire structure. Each week presents a motivational video with paired readings from the handbook and the DIME Wiki, a detailed lecture and Q&A session with DIME

Analytics team members on the topic of the week, and a knowledge assessment for attendees. The course will be offered annually, and all course materials are publicly available for self-study (DIME Analytics 2021a).

Manage Successful Impact Evaluation Surveys

Manage Successful Impact Evaluation Surveys is a free, virtual course in which participants learn and practice the workflows involved in primary data collection. It acts as a complement to the Development Research in Practice course, providing a deep dive into the processes that are described at a high level in chapter 4 of *The DIME Analytics Data Handbook*. The course covers best practices at all stages of the survey workflow, from planning to piloting instruments and monitoring data quality once fieldwork begins. There is a strong emphasis throughout on research ethics and reproducible workflows. During the global pandemic, a special module focused on adapting surveys to remote data collection. Participants learn to plan for and prepare successful surveys, design high-quality survey instruments, effectively train surveyors (including remote training), monitor survey implementation, ensure high-quality data, and handle confidential data securely, among other topics. The course uses a combination of virtual lectures, readings, and hands-on exercises. It is offered annually, and all course materials are available online for self-study (DIME Analytics 2020a, 2021b).

Measuring Development Conference

DIME Analytics annually invites a diverse group of attendees to become part of a community of analysts interested in innovations in measurement. Measuring Development is an annual conference organized jointly by DIME, the World Bank's Development Economics Data Group, and the Center for Effective Global Action at the University of California, Berkeley. The conference focuses on data and measurement innovations across different sectors and themes. It was held virtually and was open to the public in 2020 and 2021. The focus for 2021 was "Emerging Data and Methods in Global Health Research" (a summary blog can be found in Jones, Fishman, and Reschechko 2021). Previous topics have included "Data Integration and Data Fusion" (2020), "Crisis Preparedness and Response" (2019), and "Artificial Intelligence and Economic Development" (2018).⁶

DIME Wiki

The DIME Wiki is a one-stop shop for resources on the best practices and resources across all phases of an impact evaluation (IE): design, fieldwork, data, and analysis. It focuses on practical implementation guidelines rather than theory. With over 200 content pages, the DIME Wiki is open to the public, easily searchable, and suitable for users of varying levels of expertise. The DIME Wiki is closely linked to *The DIME Analytics Data Handbook*; the handbook links to the DIME Wiki for implementation details and specific examples of the best practices it outlines.⁷

Stata Packages

If you use Stata (a statistical package many researchers use for their analysis), you may be interested in the DIME Analytics Stata software packages, *ietoolkit* and *iefieldkit*. These packages are a direct result of DIME Analytics' efforts to identify inefficiencies and common sources of error in data workflows and to create tools that routinize best practices. The *ietoolkit* package includes standardized commands for data analysis tasks that are common in DIME work. The *iefieldkit* package includes commands related to primary data

collection that create rapid, standardized, and well-documented data acquisition workflows. These statistical packages can be installed through the Statistical Software Components (SSC) archive. The source code for `ietoolkit` and `iefieldkit` is available for public review and contribution via GitHub.⁸ By using these standardized commands, users avoid repeating common mistakes and produce more efficient code as well as share common workflows for major tasks.

Data Visualization Libraries

If you are producing graphics from your data, you may be interested in the DIME Analytics visual libraries for Stata and R users (Andrade et al. 2020).⁹ The libraries contain example code for model data visualizations in an easy-to-browse format. These libraries help researchers reduce the time they spend creating professional-quality, reproducible graphs and maps. Both the Stata library and the R library are open to contributions through GitHub.¹⁰

Technical Trainings

DIME Analytics provides regular technical trainings to World Bank staff and consultants. All training materials are shared with the public through the Open Science Framework platform; self-study is encouraged, and DIME Analytics provides support to independent learners by answering questions and responding to feedback in the relevant GitHub repositories. To access an index of these trainings and browse all materials, visit the DIME Analytics profile page on the Open Science Framework website.¹¹

Research Assistant Onboarding Course

The Research Assistant Onboarding Course is designed to familiarize research assistants (or, in a public administration setting, junior analysts) with best practices for reproducible research.¹² By the end of the course's six sessions, participants have the tools and knowledge to implement these best practices and to set up a collaborative workflow for code, data sets, and research outputs. Most content is platform independent and software agnostic, though participants are expected to be familiar with statistical software. The course is offered twice yearly to World Bank staff and consultants; course materials are available to the public (DIME Analytics 2020b).

R for Advanced Stata Users

The R for Advanced Stata Users course provides an introduction to the R programming language, building on knowledge of Stata.¹³ The course focuses on common tasks in development research related to descriptive analysis, data visualization, data processing, and geospatial data work. The course is offered twice yearly to World Bank staff and consultants; course materials are available to the public (DIME Analytics 2019).

Continuing Education Series

DIME Analytics offers a biweekly Continuing Education Series in the fall and spring semesters. The trainings are typically hands-on, and the topics are decided based on a review of common issues faced by DIME project teams. For example, in 2020, DIME Analytics delivered 11 Continuing Education sessions on topics including “Data Quality Monitoring,” “Working with Spatial Data in Stata,” “Optimizing Survey Length,” “GitHub Pull Requests,” and “Introduction to Python for Stata Users” (DIME Analytics 2020–2023).

CONCLUSION

The World Bank's DIME Analytics team has made a sustained, years-long effort to implement the principles of transparency, reproducibility, and credibility across the cycle of data work. It has taken an iterative approach. First, through direct engagement with analytical teams, DIME Analytics identifies processes that are both common and ad hoc, such that the whole organization would benefit from standardization. Then, the DIME Analytics team works to understand the essential needs of the analytics teams for the process or subtask, mapping each process as part of a research flowchart and defining the inputs and outputs that are desired for the process. Next, the team either identifies an external tool or process that can be utilized or develops its own tools or guidelines that are appropriate to data work. Finally, the materials are documented, archived, and disseminated into team workflows through the frequent training and support sessions the team organizes.

In this way, DIME as an organization has accumulated a knowledge base of high-quality analytical tools and standardized processes and best practices that are used across almost all its projects. For the DIME Department, these resources form a visible foundation that makes DIME research well known for its attention to quality and underscores the reliability of DIME research products. Additionally, an essential element of the mission of DIME and the World Bank is to produce global public resources for high-quality analytical evidence. In response to this mission, DIME Analytics makes its tools, processes, and trainings publicly available whenever possible and enables them to be self-paced and remote to the greatest practical extent. This chapter has summarized some of the key features and philosophies of the DIME Analytics approach and has offered these resources to readers so they may use and access whatever materials are helpful to them.

NOTES

We acknowledge the initiative of Arianna Legovini (alegovini@worldbank.org) and Florence Kondylis (fkondylis@worldbank.org) for their creation of and ongoing support for the DIME Analytics team, and Vincenzo di Maro (dimaro@worldbank.org) for his leadership and management. We especially acknowledge the work of Kristoffer Bjärkefur (kbjarkefur@worldbank.org), Luíza Cardoso de Andrade (lcardoso@worldbank.org), and Roshni Khincha (rkhincha@worldbank.org) as members of that team, as well as the support and assistance of Avnish Singh (asingh42@worldbank.org), Patricia Paskov (ppaskov@worldbank.org), Radhika Kaul (rkaul1@worldbank.org), Mizuhiro Suzuki (msuzuki1@worldbank.org), Yifan Powers (ypowers@worldbank.org), and Maria Arnal Canudo (marnalcanudo@worldbank.org). We further recognize all members of the DIME team, past and present, for their contributions to this portfolio and all the World Bank teams and external support we have graciously received. DIME Analytics has been financially supported by the United Kingdom Foreign, Commonwealth and Development Office (FCDO) through the DIME i2i Umbrella Facility for Impact Evaluation at the World Bank.

1. *The DIME Analytics Data Handbook* GitHub repository can be found at <https://worldbank.github.io/dime-data-handbook/>.
2. The DIME Wiki is maintained by DIME Analytics and can be found at <https://dimewiki.worldbank.org/>.
3. More information about the Manage Successful Impact Evaluation Surveys course is available on the World Bank website at <https://www.worldbank.org/en/events/2021/05/24/manage-successful-impact-evaluation-surveys>.
4. DIME Wiki, s.v. "Primary Data Collection," last modified May 6, 2022, 17:02, https://dimewiki.worldbank.org/Primary_Data_Collection; DIME Wiki, s.v. "Field Surveys," last modified May 6, 2021, 17:27, https://dimewiki.worldbank.org/Field_Surveys; DIME Wiki, s.v. "Survey Protocols," last modified July 11, 2022, 16:18, https://dimewiki.worldbank.org/Survey_Protocols; DIME Wiki, s.v. "Remote Surveys," last modified April 7, 2022, 17:35, https://dimewiki.worldbank.org/Remote_Surveys.
5. More information about the Development Research in Practice course is available on the World Bank website at <https://www.worldbank.org/en/events/2021/07/12/development-research-in-practice>.
6. More information about the 2018 and 2019 conferences can be found on the World Bank website at <https://www.worldbank.org/en/events/2018/01/29/artificial-intelligence-for-economic-development> and <https://www.worldbank.org/en/events/2019/03/27/crisis-preparedness-and-response>. More information about the 2020 and 2021 conferences can be found on the website of the Center for Effective Global Action at <https://cega.berkeley.edu/event/annual-conference-on-measuring-development-vi-data-integration-and-data-fusion/> and <https://cega.berkeley.edu/event/measuring-development-2021-emerging-data-and-methods-in-global-health-research/>.

7. The DIME Wiki can be found at <https://dimewiki.worldbank.org/>.
8. The GitHub repository for ietoolkit is available in the World Bank's repository at <https://github.com/worldbank/ietoolkit>. The GitHub repository for iefieldkit is available in the World Bank's repository at <https://github.com/worldbank/iefieldkit>.
9. The Stata visual library is available at <https://worldbank.github.io/stata-visual-library/>. The R visual library is available at <https://worldbank.github.io/r-econ-visual-library/>.
10. The GitHub repository for the Stata visual library is available in the World Bank's repository at <https://github.com/worldbank/stata-visual-library>. The GitHub repository for the R visual library is available on the World Bank's repository at <https://github.com/worldbank/r-econ-visual-library>.
11. The DIME Analytics profile page on the Open Science Framework website can be found at <https://osf.io/wzjtk>.
12. More information about the Research Assistant Onboarding Course is available at <https://www.worldbank.org/en/events/2020/01/30/ra-onboarding-course>.
13. More information about the R for Advanced Stata Users course can be found on the World Bank website at <https://www.worldbank.org/en/events/2019/04/16/R-for-advanced-stata-Users>.

REFERENCES

- Andrade, Luíza, Maria Jones, Florence Kondylis, and Luis Eduardo (Luise) San Martin. 2020. "New Visual Libraries for R and Stata Users." *Development Impact* (blog). *World Bank Blogs*, October 15, 2020. <https://blogs.worldbank.org/impacitevaluations/new-visual-libraries-r-and-stata-users>.
- 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>.
- Dabalen, Andrew, Alvin Etang, Johannes Hoogeveen, Elvis Mushi, Youdi Schipper, and Johannes von Engelhardt. 2016. *Mobile Phone Panel Surveys in Developing Countries A Practical Guide for Microdata Collection*. Directions in Development. Washington, DC: World Bank. <http://documents.worldbank.org/curated/en/877231468391801912/Mobile-phone-panel-surveys-in-developing-countries-a-practical-guide-for-microdata-collection>.
- DIME Analytics. 2019. "R for Advanced Stata Users." Recordings and Course Materials from Virtual Course Held April 16, 2019, World Bank, Washington, DC. Created March 17, 2020, and last updated May 31, 2022. Open Science Framework. <https://osf.io/nj6bf/>.
- DIME Analytics. 2020a. "Manage Successful Impact Evaluation Surveys." Recordings and Course Materials from Virtual Course Held July 6–24, 2020, World Bank, Washington, DC. Created June 3, 2020, and last updated May 31, 2022. Open Science Framework. <https://osf.io/resya/>.
- DIME Analytics. 2020b. "Research Assistant Onboarding Course." Recordings and Course Materials from Virtual Course Held January 30–March 5, 2020, World Bank, Washington, DC. Created March 31, 2020, and last updated May 31, 2022. Open Science Framework. <https://osf.io/qtmdp/>.
- DIME Analytics. 2020–23. "DIME Continuing Education." Recordings and Course Materials for Biweekly Virtual Course Series, World Bank, Washington, DC. Created March 20, 2020, and last updated February 17, 2023. Open Science Framework. <https://osf.io/8sgrh/>.
- DIME Analytics. 2021a. "Development Research in Practice." Recordings and Course Materials from Virtual Course Held July 12–September 3, 2021, World Bank, Washington, DC. Created June 24, 2021, and last updated May 31, 2022. Open Science Framework. <https://osf.io/6fsz3/>.
- DIME Analytics. 2021b. "Manage Successful Impact Evaluation Surveys." Recordings and Course Materials from Virtual Course Held May 24–June 11, 2021, World Bank, Washington, DC. Created April 30, 2021, and last updated May 31, 2022. Open Science Framework. <https://osf.io/672ej/>.
- Jones, Maria, Sam Fishman, and Yevanit Reschechko. 2021. "Data-Driven Global Health Research in the Time of COVID." *Data Blog*. *World Bank Blogs*, April 26, 2021. <https://blogs.worldbank.org/opendata/data-driven-global-health-research-time-covid>.

CHAPTER 6

The Ethics of Measuring Public Administration

Annabelle Wittels

SUMMARY

As data collection *on* government *within* government becomes more prevalent, a review of research on data ethics fit for use within public administration is needed. While guides on data ethics exist for public sector employees, as well as guides on the use of data about citizens, there is a dearth of discussion and few practical guides on the ethics of data collection by governments about their own employees. When collecting data about their employees, public administrations face ethical considerations that balance three dimensions: an individual dimension, a group dimension, and a public-facing dimension. The individual dimension comprises demands for dignity and privacy. The group dimension allows for voice and dissent. The public-facing dimension ensures that analytics enable public servants to deliver on public sector values: accountability, productivity, and innovation. The chapter uses this heuristic to investigate ethical questions and provide a tool (in appendix B) with a 10-point framework for governments to guide the creation of fair and equitable data collection approaches.

ANALYTICS IN PRACTICE

- Collecting data on government employees involves a different set of challenges from collecting data on service users and private sector employees.
- Depriving public sector employees of privacy can erode democratic principles because employees may lose spaces to dissent and counteract malpractice pursued by powerful colleagues, managers, or political principals.
- Designing data collection in a way that enhances dignity serves several functions. For one, if we understand the problem of privacy in terms of disruptions of autonomy, collecting data in ways that enhance dignity can help to move away from zero-sum thinking. If public sector employees gain dignity from data collection, they are unlikely to see it as an unwanted intrusion. For instance, data could be collected that celebrate personal initiative, good management practices, and outstanding achievements, as opposed to disciplinary uses of data, such as identifying malpractice or inefficiencies. In some cases, employers

Annabelle Wittels is an independent researcher and former consultant in the World Bank's Development Impact Evaluation (DIME) Department.

might also consider disclosing the identities of participating individuals, if they consent, to highlight valuable contributions and give credit.

- Despite improved data quality, which helps to produce efficiencies, value-based decision-making will not become obsolete. Negotiating values is required as much as ever to produce evidence-based and ethically sound policy (Athey 2017).
- Development practitioners and donor countries working on data strategies for public sector reform in countries where political, religious, or other basic human freedoms are not guaranteed must thus tread carefully to guard against setting up data infrastructures that can be used to the detriment of public sector employees.
- Navigating ethical dilemmas sustainably requires that, when individuals join social groups in which different norms on data privacy are applied, they do so knowingly and are provided with the opportunity to advocate to change these norms. In practice, this can mean giving public sector staff unions a voice in what data are made available about their members and what guarantees of accountability they can offer the public in place of such data.
- Creating data approaches for public sector innovation thus requires that time and resources be set aside to make the process explicable to those affected. This is no simple task because governments still face skills gaps in cutting-edge areas of information technology. In many instances, governments will need to rely on external expertise to develop and maintain the skills of their staff to implement data solutions that are ethically sound and secure.
- What is considered ethical and morally right can depend on context. There are, however, questions that provide general guidance for how measurement can be conducted in an ethically sound manner, if they are asked regularly at key junctures of data collection, analysis, and reporting.
- It is important to construct objective measures of organizational and/or individual performance rather than relying only on subjective evaluations, such as performance appraisals.
- To construct an objective measure of performance using case data, one should ensure that cases are comparable to one another. This could entail comparing cases only within a homogeneous category or constructing a metric that captures the complexity of the case.
- Measures of performance for public sector organizations will depend on the specific context of study and data availability, but they should reflect both the *volume* of services provided as well as their *quality*.

INTRODUCTION

Data, from the Latin *datum*, meaning “given,” are the embodiment of something factual, technical, value-free. Yet, data are more than that. They have and create monetary worth: “Data is the new capital” has become the catchphrase of the 21st century as their importance in value generation has increased. Data are central to the business models of most leading Fortune 500 companies (MIT Technology Review and Oracle 2016; Wang 2012). Their potential for poverty alleviation, growth, and development has been recognized. For instance, the World Bank’s *Europe and Central Asia Economic Update, Spring 2021: Data, Digitalization, and Governance* places “data” center stage (World Bank 2021a). Several governments have already demonstrated how they can use data to provide better services and protections to their citizens (UN Data Revolution Group 2014; World Bank 2021a). Collecting data on people who work in government has become part of using metrics to improve government service delivery (see table 6.1 for an overview of the types of data that governments can collect on their workforce and examples of how they can contribute to a mission of service improvement).

TABLE 6.1 Types of Data Collected on Public Administration, with Examples

Type of data collected on public sector employees	Examples
Prehiring metrics	Qualifications, work experience, gender, age, ethnicity, sexual orientation, disability
Recruitment metrics	Test and application scores, background checks
Performance metrics	Output rate, user feedback, supervisor reviews
Learning and development metrics	Rate of promotion, courses taken
Incentives	Disclosing salaries, disclosing tax returns, pay scales, bonuses
Survey data	Attitudes, self-reported behaviors, practices
Linked data	Survey + administrative data, survey + geospatial data, administrative + geospatial data

Source: Original table for this publication.

Parallel to increased awareness of data as a source of value creation, greater attention is being paid to how rendering observations into data relies on structures of power and value trade-offs. Over the last decade, public debate on the ethics of data use has become a fixture of global news, with most articles focusing on the use of consumer data (Bala 2021; BBC News 2019; Fung 2019; Pamuk and Lewis 2021), some on governments’ use of data on citizens (for example, Beioley 2022a; Williams and Mao 2022; World Bank 2021a), and some on companies’ use of data on their employees (Beioley 2022b; Clark 2021; Hunter 2021; Reuters 2022). The intersection of the last two arenas—the use of data by governments on their own employees—has received next to no attention. This is likely because governments are only starting to catch up in the use of employment and within-government metrics and because claims to privacy are more complicated in the case of government employees.

This chapter tries to address this gap by providing a thorough, albeit not exhaustive, discussion of ethical issues specific to data collection on government employees. It regards *data* as a multifaceted construct: an amalgamation of points of information that—depending on how they are processed and analyzed—can become capital, commodity, truth, or all at once. Data’s function as truth, in particular, distinguishes them from other resources that have shaped world economies over the last centuries. Neither gold, oil, nor 5G has inherent value because of what it says about human behavior and the world we live in. In that sense, data and their use raise ethical conundrums not seen before in other phases of technological adoption.

Data linkage and triangulation offer the best chance for constructing measures of public sector productivity that are meaningful and provide an acceptable level of accuracy. Such developments have brought their own problems. Greater triangulation can lead to greater reliance on indicators. See the discussion in chapter 4 of this *Handbook*.

Choosing these indicators means choosing what to make salient in an official’s work environment.

Guides to ethical data use for both the private and public sectors now exist in many jurisdictions (Mehr 2017; Morley et al. 2020a; OECD 2019; Office for Science 2020; Open Data Institute 2021). Guidelines for the public sector tend to be heavily modeled on those for the private sector, meaning that they mostly speak to data interactions between the government and public service users. There is, however, a significant amount of data concerning processes and people within government. Collecting data on government employees involves a different set of challenges from collecting data on service users and private sector employees (see table 6.1 for an overview). The ethics of collecting data on government employees thus merits a separate discussion.

This chapter uses the terms “government employee,” “public sector employee,” and “civil servant” interchangeably. The majority of the concerns discussed here are relevant to people employed in central government functions as much as those employed in frontline services (“street-level bureaucrats”). The proportionate relevance of concerns will differ by the type of public sector employment and the regime type. As the aim here is to provide a general framework for data collection on employees in the public sector,

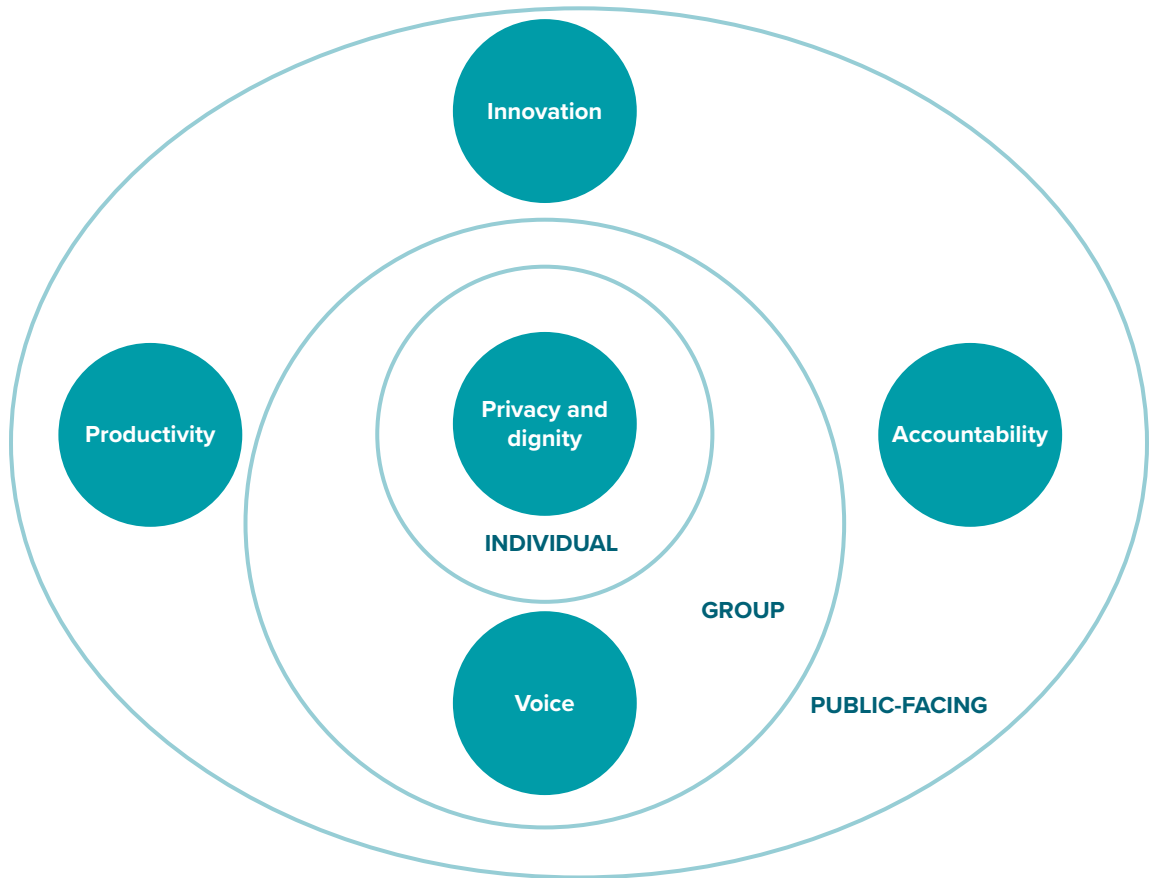
differences will not be discussed in detail, in favor of providing more space for discussion of how ethical data collection can be put into practice in the public sector.

Data and Public Value

What public value *is* in and of itself subject to debate. Seminal scholarship on the topic, like Mark Moore’s work on public value (Moore 1995), eschews defining public value by ascribing the task to public managers and their interactions—perhaps better called “negotiations”—with political and societal agents (Rhodes and Wanna 2007). Deborah Stone’s (2002) work on public value highlights the ideals of equality, liberty, and security core to the idea of statehood and the role of government but stresses that none of these ideals can be fully catered to in most policy settings.

For a working definition of public value that serves the debate about the use of data on government employees, this chapter will focus on three broad aspects of public sector endeavors that can produce public value by producing goods and supporting equitable information exchange between the governed and the governing: accountability, productivity, and innovation (see figure 6.1).

FIGURE 6.1 Ethical Dimensions That Require Balancing in Data Collection Efforts on Public Sector Employees



Source: Original figure for this publication.

Note: As a heuristic, one can imagine key questions about the ethics of collecting data on public sector employees falling into three circles. The innermost circle describes the individual dimension. These questions mainly concern the privacy and dignity of the individual public sector employee. The middle circle signifies the group dimension. These questions concern voice and dissent, which are central to many functions that public sector employees carry out and the tensions that arise when collecting data on them as groups. The third, outermost circle encapsulates questions related to the qualities that define public sector value creation in relation to stakeholders: political principals, public service users, and society at large.

The chapter discusses accountability first because it is central to ethical exchanges between citizens and governments and employers and employees. It next focuses on productivity because it considers the creation of public value in the sense of government output. Finally, it turns to innovation because concerns about public value creation are not limited to what is produced but, particularly in the context of big data and artificial intelligence, include whether data used for public value creation can also produce innovation in the form of improvements or altogether new outputs. The chapter discusses all three concepts with a view to the wide-ranging types of data that can be collected to inform public administration, such as pre-hiring, recruitment, performance, and learning and development metrics (see table 6.1).

The Three Dimensions of Public Sector Data Ethics

As a heuristic, the ethical considerations facing public administrations when collecting data on their employees comprise three dimensions: an individual dimension, which comprises demands for dignity and privacy; a group dimension, which relates to voice and dissent; and a public-facing dimension, which ensures that data enable public administrators to deliver on public sector values.

The Individual Dimension: Individual Dignity and Privacy

Individual demands for dignity and privacy—the first dimension of ethical concern for employees—have been discussed widely in standard volumes on research ethics (Carpenter 2017; Macdonald 2017). Although the other two dimensions have received less attention, the first still merits discussion because, owing to their unique position in political-economic systems, public sector employees face a different set of demands, incentives, threats, and opportunities than private sector employees. For instance, because of the public sector's dominance in providing jobs and offering better wages in many countries (Düwell et al. 2014), exit options for public sector employees are more limited than for those employed in the private or nonprofit sector. This has implications for informed consent. Employees might accept trade-offs because of the constrained choice sets they face rather than satisfying their needs for privacy and dignity.

The temporal and spatial dimensions of how data on employees are safe-guarded also differ from the private sector. What is considered “good public service” and service “in the national interest”—and thus what types of measurement and data use are justified—can shift as governments and their view of the nation-state and societal ideals change. Such shifts in value positions affect both individual freedoms and those of groups within the public sector.

Another difference is created by the pressure to reflect political shifts in how government-owned organizations are run. For example, state-owned enterprises not only deliver services where market failures exist but also serve as model employers (for example, by pioneering inclusive hiring practices), act as symbols of national identity (for example, airlines and national health services), and can play a redistributive function (for example, by providing subsidized goods or privileged access to select groups of citizens and stakeholders; see Heath and Norman [2004] for a more extensive discussion of state-owned enterprises). The aim of data collection on individuals and groups in the public sector might thus change over time compared with the private sector. Some political factions believe many services and functions should not be performed by the public sector at all, or if they do not go this far, they have a deep-seated mistrust of civil servants. The threat of surveillance and a push to replace workers with compliant and efficient machines thus might be even more acute for the public sector than the private sector, depending on the political leaning of the government in power. As a case in point, remote and flexible work has become standard in many industries. Because of competition among companies and sometimes even industries (think aerospace engineers joining tech companies or banks), these standards are unlikely to be reversed. Data on sick days and other types of leave taken by employees in the United Kingdom suggest that private and public sector workers are absent for fewer days in a year than public sector workers. These data, and the supposedly empty seats (due to remote work) in government halls, led leaders of the UK Conservative Party to campaign for remote-work policies to be curtailed

(BBC News 2021; Cabinet Office 2022; Lynn 2022; Office for National Statistics 2022). How data on public sector workers are collected and used might change more with the political flavor of the day than with the industry standard or best practice in other fields.

The Group Dimension: Voice and Representation

The second, group dimension becomes relevant in a setting where data relate to existing or ad hoc groups of employees. The group dimension recognizes that employees as a group have a right to actively shape what is done to them: they have a right to exercise *voice*.

Hirschman (1970) first introduced the concept of *voice*, alongside *exit* and *loyalty*, to define the dynamics of engagement with institutions. *Voice*, in this sense, is defined as the expression of opinions and thoughts in a manner that is impactful or at least has the potential to be impactful. Denying stakeholders the option of voice makes exit more likely, which, in an organizational setting, means disengaging from the cause or leaving the organization. Voice is thus a moral imperative as much as a practical necessity.

The need for voice in public service creates many ethical conundrums. Voice is necessary because “freedom defined as noninterference or the absence of burdensome restraints clearly will not do” (Preston 1987, 776). Civil servants need space to speak up, take initiative, disagree, and innovate.

On the other hand, for organizational goals to be attained, the expression of voice needs to be bounded. It requires agreement on a narrative about what is good, achievable, and desirable. This is particularly true in many public service organizations, where adherence to mission is an important guide to action and a motivator for public servants as a whole. Voice may place individuals or groups in the position of identifying themselves as in defiance of, or as distinct from, the prevailing public service culture.

Boundaries to voice are also necessary because of the mechanics of representative democracy: central to the role of civil servants is that they implement directives from democratically elected or appointed political leaders. Civil servants thus need to subordinate some of their own concerns to the policies elected representatives choose. Such subordination can be demanded more easily of individual civil servants. However, when voice is exercised on behalf of groups—for example, women, ethnic minorities, and people with disabilities in the civil service—boundaries are much more difficult to draw. Civil servant groups, then, are both implementers of democratic will and constituent groups with a right to voice at the same time.

The Public-Facing Dimension: Data to Operate in Service of the Public

The third, public-facing dimension is particular to the public sector because it concerns data collected with a view to serving public service functions. It highlights the ethical challenges linked to creating public value and the organizational capabilities required to do so (Panagiotopoulos, Klievink, and Cordella 2019). Data collection and use must be designed in a way that enables government employees to operate effectively and efficiently, to collaborate and innovate.

Challenges located within this third dimension include balancing data security requirements, openness for innovation, and room for experimentation, as well as the certainty, consistency, accountability, and reliability of public service. While not necessarily mutually exclusive, these demands create tricky trade-offs. For instance, making data on what public sector workers do available can help the population to monitor them and call out malpractice; it can also help others spot ways of innovating, doing things better or differently; however, it can also create fear and political suppression and encourage inflexibility.

The following sections will discuss each ethical dimension in turn. Each section first outlines specific demands and incentives facing public sector employees before discussing how data can help to address ethical challenges and where they introduce new challenges that require closer scrutiny. Appendix B includes a framework for evaluating the ethics of measuring and tracking public sector workers. Practitioners can use this framework to think through the key tensions laid out in this chapter.

INDIVIDUAL LEVEL: DIGNITY AND PRIVACY OF PUBLIC SECTOR EMPLOYEES

Dignity and, by extension, privacy are two values central to the ethical handling of data. Dignity (Schroeder and Bani-Sadr 2017) and privacy (Solove 2008, 1–25) are concepts that have caused controversy because they are complex and, at times, tautological. This chapter employs *dignity* to mean that humans—in this case, public sector employees—have worth in and of themselves that *they themselves* can define. This stands in contrast to seeing people as *means* to effect government outputs and *ascribing to them the essence of their ends*. Practically, dignity requires respect for the individual as a human being with free will, thoughts, and feelings. This means that the options available to people on whom data are collected should be central to the design of research efforts. For instance, if the designers of employee web pages decide it is nice to show pictures of staff, individual employees should be allowed to have their pictures removed if they do not want others to know what they look like, regardless of whether the employer agrees that this concern is central to what it means to respect their employees.¹

Privacy, as it is used in this chapter, does not mean *anonymity*. Information might be disclosed anonymously—for instance, via a survey where no names are provided and IP addresses or similar identifying characteristics of the respondent cannot be linked to responses. However, this act still entails giving up privacy because what was internal and possibly unknown to others is now known to others. The reasons why such a strict definition of privacy is adopted in this chapter become clear when discussing the group-level dimension of data ethics concerning public sector employees: even when information cannot be linked to an individual, as soon as information can be linked to a group—public sector employees—their thoughts, behaviors, and environs become known to people other than the members of this group. Discussions of *individual* privacy, the focal point of this section, must therefore be separated from discussions of *collective* privacy, which will appear in later sections.

Relatedly, *privacy* as it is used here is understood in terms of a “typology of disruptions” (Solove 2008), which acknowledges that the definition of privacy is highly contextual. Just as quirks are distinguished from mental health disorders, disclosure, transparency, and openness are distinguished from infringements on privacy by the disruption they cause. Does a person only occasionally need to return home to check whether the stove was left on, or is this a daily occurrence that interferes with a person’s life? Is it a minor issue that a public servant’s address is publicly available online, since everyone’s address is publicly available online, or is it a danger to the public servant’s safety and right to conduct personal matters in private? Until the 1980s and into the late 1990s, it was common in Western European countries for the phone numbers of the inhabitants of entire cities to be listed, publicly available in white pages or equivalent phone and address directories. In several Scandinavian countries, it is still the case that every resident’s social security number, address, and taxable income is made publicly available. The key differences are the extent to which something is experienced as a disruption, as opposed to the norm, and the extent to which people can stop a practice if they start to experience it as a disruption. In the case of telephone and address registries in democracies, residents can use their voting powers to change the laws surrounding the publication of personal details. It is less clear how employees—particularly state employees, of whom transparency is expected—can demand change when they find practices intrusive. Privacy, as defined in this chapter, is thus closely linked to the idea of control: the extent to which civil servants control how much is known about them when they experience it as intruding on realms they perceive as reserved for their private as opposed to their professional (work) persona.

Commonly, informed consent procedures serve to preserve dignity by affording individuals the opportunity to ascertain what they see as acceptable in how they or their data are handled. Informed consent means that the individuals on whom data are collected or who are asked to divulge information are fully informed about the purpose of the research, how their data are handled, and how they will be used. Even if they agreed at an earlier stage, individuals ought to be given the right to withdraw their consent at any stage, which requires the secure deletion of any data collected on them. Typically, there are few options for public officials to opt out of servicewide databases.

The extent to which informed consent is de facto voluntary is important. In situations of near employer monopoly, in which exiting the public sector is not a viable option, and in situations of suppression, in which exercising voice is not possible, employees might consent to data collection because they see little other choice.² This mirrors problems with consent ubiquitous in today's highly oligopolistic landscape of online service providers: if one is pressed to quickly find a parking spot and only Google Maps can pinpoint one with timeliness and accuracy, one is likely to accept the data use conditions that come with this service. The viability of informed consent is intricately bound to the ability to exercise exit and voice.

Valuing dignity also extends to claims to privacy. People ought to be allowed to keep certain details of their lives and personhood protected from the scrutiny of others—even their human resources (HR) managers. Any form of data collection on individuals will require some abnegation of privacy. Research ethics typically tries to acknowledge this right by providing confidentiality: a declaration that the information disclosed to the data collector will be known to and used by a limited, preidentified set of people. In the public sector, this is the case, for example, when an internal research or strategy unit collects survey responses from civil servants with a view to sharing them with HR and finance operations to improve planning, staffing, and training.

The principle of confidentiality in research brings to mind that research ethics has its foundation in the Hippocratic Oath and in bioethics. Patients trust their doctors to share their medical information and concerns only with relevant personnel, and always with the intention to help the patient. In a medical setting and in most everyday circumstances, we tend to assume confidentiality and do not give explicit consent to use otherwise private details divulged as a part of interactions. We do so willingly because disclosing information encourages reciprocity, builds trust, and helps shape the world around us to better meet our needs. Reductions in privacy are not a zero-sum game but can offer substantial welfare gains for the individual and society at large when negotiated carefully. The individual is, however, poorly positioned to negotiate these tradeoffs against the interests of large corporations or the government. As discussed above, this is particularly true if an individual would like to exercise exit or voice when the options presented to them do not inspire trust.

In response to this problem, legal protections have been put in place to guard against the worst misuses of data. Data regulations such as the European Union's General Data Protection Regulation (GDPR) require that entities collecting data clearly lay out which data are collected, how they are handled, and who has access to them. California, Virginia, New Zealand, Brazil, India, Singapore, and Thailand have all implemented legislation similar to the GDPR in recent years. However, detailing how data are used typically leads to documents that require, on average, 10 minutes to read (Madrigal 2012). As data collection has become ubiquitous, the time burden that consent processes introduce implies that most of us have become accustomed to quickly clicking through consent forms, terms and conditions, and other common digital consent procedures. This means that in practice, consent is either based on trust rather than complete knowledge or, in the face of a lack of exit and voice options, is coerced.

Following legal guidelines is thus not enough to ensure that dignity and privacy concerns are adequately addressed. Those planning to collect data on public sector employees must take into account what ethical challenges could arise, how to offer exit and voice options, and how to foster trust. This is not a simple feat. The discussion will thus next turn to how three common dilemmas concerning data privacy and dignity in the public sector can be addressed.

Government Employee Data, Dignity, and Privacy

Discussions of data ethics concerning the dignity and privacy of civil servants could fill volumes. This chapter, therefore, cannot offer a comprehensive account of the debate. Instead, it focuses on three likely areas of concern for someone wanting to collect data on civil servants: trade-offs surrounding access, transparency, and privacy; how data collection can be designed to enhance dignity; and how data transfer and storage should be managed to safeguard privacy and dignity.

Trade-Offs Surrounding Access, Transparency, and Privacy

When collecting data on public sector employees, it can be argued that knowing about their behaviors and attitudes is in the public interest. For example, using data to identify inefficiencies associated with public employees is only possible without their active consent, but is certainly in the public interest. In many jurisdictions, public sector workers do not have to be asked for their consent for research to be conducted on them as long as it can be shown that the research is in the public interest. In most countries, where these provisions are not made explicit in the law, data ethics codes include clauses that allow the requirement for consent to be lifted if there is a strong case that research is in the public interest. These waiver clauses tend to use unequal power relations as grounds for the lifting of consent requirements: researchers might be prevented from gaining access to public institutions if they require explicit consent from people in positions of power who are hostile to the idea of research (Israel 2015). Based on the public-interest argument, consent procedures can, in many instances, be circumvented when researching public sector employees.

A reduction in privacy and an overruling of informed consent can thus promote *accountability*. They can also enhance *transparency*. Having more in-depth knowledge of the behavior of government employees can help to increase government transparency: just as the work of election volunteers was live streamed during the 2020 US elections (Basu 2020), civil servants could be put under constant surveillance to increase transparency and inspire trust. Evidence from the consumer context suggests that extreme levels of transparency can create a win-win situation, in which customers rate services more highly and are willing to pay more when they can monitor how a product is created or a service delivered (Buell, Kim, and Tsay 2017; Mohan, Buell, and John 2020). Radical transparency could thus inspire greater trust and mutual respect between government employees and government service users as opposed to simply reducing dignity and privacy by increasing surveillance.

However, promoting one type of public good might infringe on another (see the discussions of contradicting policy choices in Stone 2002). Privacy is instrumental to guarding collective freedom of speech and association (Regan 2000). Depriving public sector employees of privacy can erode democratic principles because employees may lose spaces to dissent and counteract malpractice pursued by powerful colleagues, managers, or political principals. To date, the ambiguity of public-interest claims has been most commonly revealed in cases of whistleblowing (Boot 2020; Wright 1987): government whistleblowers often endanger some public interests (for example, national security) in favor of others (for example, transparency). How convoluted claims to public interest can become is highlighted when whistleblowers reveal previously private information about some actors with a public-interest claim (for example, disclosing that a particular person was at a particular location at a particular time or making private communication between individuals public). The privacy of whistleblowers needs to be protected in order to shelter them from unfair prosecution and attacks so that future whistleblowers do not shy away from going public. The future public interest, then, is guarded by protecting the present privacy of the whistleblower, who might have rendered public what was previously thought to be private “in the public interest.”

In this sense, what is in the public interest and what the limits are to a utilitarian logic of increased surveillance must remain part of the public debate. For public debate to be a viable strategy for dealing with the contradictions of disclosure and the protection of privacy in the public interest, society must protect fora and institutions that publicize such issues, involve powerful stakeholders, and have tools at their disposal to enforce recommendations. To date, this often means supporting the capacities of the media, civil society, the political opposition, and the judiciary to fulfill these functions. Researchers and governments thinking about collecting data on government employees need to assess whether these institutions function sufficiently or, if not, whether actions can be taken to remedy their absence. For instance, governments could create independent review committees, actively publicize planned data collection efforts, and provide extra time for public consultations. In settings where such mechanisms lack potency, the international community, most likely in the form of intergovernmental organizations and donors, has a responsibility to monitor how changes in data regimes affect the dignity, privacy, and welfare of data subjects and the citizenry more broadly.

This is important not solely with a view to balancing trade-offs between transparency, public access, and privacy. The next section looks at how a thorough design and review of government data collection strategies could help to create regimes that enhance dignity despite entailing reductions in privacy.

Designing Data Collection to Enhance Dignity

Designing data collection in a way that enhances dignity serves several functions. For one, if we understand the problem of privacy in terms of disruptions of autonomy, collecting data in ways that enhance dignity can help to move away from zero-sum thinking. If public sector employees gain dignity from data collection, they are unlikely to see it as an unwanted intrusion. For instance, data could be collected that celebrate personal initiative, good management practices, and outstanding achievements, as opposed to disciplinary uses of data, such as identifying malpractice or inefficiencies. In some cases, employers might also consider disclosing the identities of participating individuals, if they consent, to highlight valuable contributions and give credit. (See also Israel [2015] for case studies of how sociological and anthropological researchers typically acknowledge the role that research participants play in scientific discovery.)

Apart from a normative view of why data collection efforts should enhance dignity, there are clear utilitarian reasons: this can improve data quality, trust, and openness to data collection and data-based management practices. Public sector employees might refuse to disclose their true opinions when they feel pressured into consenting to data collection. This can, for instance, be the case when managers or political principals provide consent on behalf of their employees. Lifting consent procedures with public-interest claims can backfire in such cases. Engaging with employee representatives and living up to promises of transparency concerning the objective of data collection can help.

Processes that ensure that data collection is linked to clear action points can further help to guard the dignity of research participants. Data collectors have an additional incentive to do so because the validity of responses and response rates will likely deteriorate when staff see that data collected on them are not used to their benefit.

Staff will more likely develop trust in the process and engagement with results when they have a stake in developing research questions and action plans following up on results. Principles of action research, including building phases of review, consultation, and revision into the research process, could help to create data collection strategies on public sector employees that enhance the dignity of the individuals involved.³

Designing Data Transfer and Storage to Guard Dignity and Privacy

Both dignity and privacy are at risk when data are not secure. Cyberattacks are becoming more common. For example, in 2021, the Washington, DC, Police Department was subject to a massive data leak following a ransomware attack. Disciplinary files and intelligence reports including names, addresses, and sensitive details about conduct were leaked into the public domain (Suderman 2021). In 2015, the Office of Personnel Management (OPM) of the US federal government was subject to a hack that led to the leaking of the personal data of millions of employees, many of whom suffered from identity theft for years following the data breach (CBS and the Associated Press 2015).

Guarding dignity and privacy in this sense is as much a technical as a moral issue. Legal frameworks such as the GDPR have been created with this in mind. Several international best-practice guides on data protection elaborate on the technicalities of such efforts. Good examples include sections on data protection in *Development Research in Practice: The DIME Analytics Data Handbook* (Bjærkefur et al. 2021) and the Organisation for Economic Co-operation and Development's (OECD) Privacy Guidelines (OECD 2022).

GROUP LEVEL: VOICE AND DISSENT

As the chapter so far has reviewed, there are many aspects of data collection that affect government employees as individuals. These are perhaps most comparable with the concerns affecting private persons and research subjects. There is, however, a dimension that becomes particularly important when thinking of public sector employees as a group and the groups within the public sector that can be created based on observable characteristics or self-elected labels.

As described above, the concept of *voice*—alongside *exit* and *loyalty*—was coined by Hirschman (1970). In the public sector, voice amounts to the ability of employees to express opinions and the potential of these expressions to impact how public administrations are run. When exit options are limited, voice becomes a more pertinent tool for employees to exercise control over their environment. In public sector employment, voice is also conceptualized as part of the job of civil servants. For instance, the UK's Civil Service Code demands that public administrators “provide information and advice, including advice to ministers, on the basis of the evidence, and accurately present the options and facts” and that they “not ignore inconvenient facts or relevant considerations when providing advice or making decisions” (UK Government 2015). Voice in this function is mainly intended to enable elected officials to deliver on their policy programs. If, however, elected officials endanger public sector values and the integrity of rule-based and meritocratic government institutions, a professionalized bureaucracy is expected to exercise voice to counteract this erosion. This can take place within legitimate reimits of voice and discretion (Miller and Whitford 2016), or it can take the form of dissent (Kenny 2019).

Demands for voice are intricately linked to those for productivity and innovation. In organizational studies, it has long been established that psychological safety—the feeling that employees can voice ideas and concerns without facing an immediate threat to their jobs or selves—is necessary for innovation and sustainable increases in organizational performance (Baer and Frese 2003; Nembhard and Edmondson 2006). Empirical research shows that more-diverse workplaces, in the private and public sector, are more-productive workplaces (Alesina and Ferrara 2005; Hjort 2014; Rasul and Rogger 2015).

Voice has also been conceptualized as civil servants' representation of the interests of the demographic groups to which they belong—a kind of “passive voice.” It is theorized that they do so through increased attention to these groups' needs and a heightened awareness of how to design and deliver public services to address them (summarized under theories of representative bureaucracy; see Kingsley [1944] 2003; Meier 2019).

Data on civil servants can help to promote or curtail voice in its active and passive forms. With regard to the passive form of voice, data can showcase how certain groups (for instance, women or bodily disabled people) are affected by or think differently about certain employment policies. With regard to its active form, data can be used by groups to raise issues themselves. For example, if the number of bullying or sexual harassment complaints filed by a department or government subdivision is made public, victims understand that their cases did not happen in isolation, and numbers can be used to pressure leadership for change.

Government Employee Data, Voice, and Dissent

Data on groups of public sector employees raise ontological and utilitarian questions. The former questions relate to how data can define groups and how groups can use data to define their workplace, work, and position between political principals and citizens.⁴ The latter questions concern how the risks and benefits of using data relate to attempts to improve working conditions and service delivery in a way that is responsive to the needs and wants of groups that become apparent through the group-level aggregation of data.

Assigning group labels to individuals implies that their individual rights and identities are pegged to those of a group—potentially one with which they do not identify or of which they did not consent to be part. Such *passive group* membership has typically been applied to individuals grouped together as “vulnerable” or “fragile” populations (Grantham 2020, 39). As the now-mainstream debate on gender and gender pronouns has raised, similar questions can be applied to group labels that have traditionally been considered more stable.

Consent can be a vehicle to ensure the alignment of grouping with personal preference. For instance, in surveys, civil servants can opt out of providing demographic details. However, for most administrative data collection efforts, consent is limited or wholly unfeasible. Employees might be grouped together as “fit for early retirement” or as a target group for “offering overseas posting” because of their age, gender, tenure, family situation, and other administrative data available to other administrators. These groups might not align with the desires or career ambitions of the grouped individuals. Basing planning decisions solely on results arrived at from demographic or ad hoc grouping risks wrong conclusions. The availability and increasing richness of data available on groups thus cannot substitute for meaningful engagement with them. These arguments are touched on by Bridges and Woolcock in chapter 4.

The creation of groups and the pegging of data to group labels not only creates risks; it also holds immense positive power. Data can open up avenues for voice because they showcase where differences between groups exist. For instance, in 2014, the cross-organizational initiative Data2X started a global drive to increase the collection and use of data on women and gender-disaggregated data. The initiative has helped to increase women's access to finance and meaningful work and has significantly reduced maternal mortality rates (Grantham 2020). In the context of collecting data on public sector employees, data can help practitioners better understand issues such as the proportionality of the representation of groups in different positions and sections (for example, Are women proportionally represented in leadership positions? Does the proportion of civil servants from minority groups map onto their proportion in the general population?); planning for skills gaps (for example, Are there enough people with advanced IT or data analysis skills in each department?); or spotting management problems in part of the civil service (for example, Do people staffing local government offices have the same level of goal clarity as those working in central locations?). In this sense, data can increase voice and benefit society.

Navigating the collection and use of data on public sector employees requires moving beyond acknowledging how data shape realities to discussing how the risks and benefits created by this process can be negotiated. Data can catalyze employees' voice by giving groups a platform to assess metrics pertaining to these groups compared with relevant benchmarks. For instance, detailed employee data including demographic details can help practitioners to understand what predicts turnover and whether certain staff members—ethnic minorities, women, or people with care responsibilities—can be retained as well as others (Grissom, Viano, and Selin 2016). Where performance data are available, data on staff can be linked in order to better understand how staffing decisions affect service delivery. For example, employing ethnic minority teachers in publicly funded schools in otherwise homogeneous districts has been associated with better performance for ethnic minority pupils attending these schools (Ford 2022).⁵ Data disaggregated by groups can also help provide better access to training and career progression for public sector employees.

As Viswanath (2020) notes, data equity, in terms of knowing what ranks people from different sections of society can attain in public service, is critical to providing better service (in line with theories of representative bureaucracy) but also to providing greater equity in opportunity for public sector staff. As a case in point, public sector unions in Canada have successfully used wage data disaggregated by gender to support calls for gender-based pay parity (Card, Lemieux, and Riddell 2020). This has created more equality between men and women in the public sector and, as a consequence of the non-negligible amount of the population employed in the public service, has improved pay equality across a large section of society.

At the same time, there is no guarantee that the availability of data will safeguard the rights of minority groups and promote equity and equality of opportunity. Data on groups can also be used to curtail voice. For example, while collecting recruitment metrics could heighten the potential for governments to hire a more diverse workforce, it could equally enable them to weed out people who are deemed less desirable. Such weeding out could be based on demographic details, but it is increasingly also founded on what potential employees say online. Hiring managers can easily introduce bias into the hiring process if the recruitment process is not sufficiently anonymized. For instance, it might be important to collect data on which universities applicants attended. These data, however, can also be used by hiring managers to prescreen candidates—consciously or unconsciously—based on associations of quality and merit with these universities. In a similar vein, even though hiring managers might not get access to detailed information on routine background checks, they can use an applicant's name and previous employer or university affiliation to conduct their own online searches.

Indeed, public sector unions in Canada and Australia now actively discourage their members from posting online or having an online presence that can be linked to their identities, in fear of potential discrimination for current employment and future employment opportunities (Cooper 2020a, 2020b). In some government contexts, there is the danger that governments collect information on employee opinions systematically. This is problematic not only at the individual but also at the group level. Investigations by big data scholars have illustrated how, for instance, people using hashtags on Twitter related to the Black Lives Matter movement could be grouped together (Taylor, Floridi, and Sloot 2016, 46). In a public sector context, such information could be used to map out the political affiliations of employees.

Other administrative data could be used for targeting based on religion or personal circumstances. For instance, data scholars have shown that people can be identified as devout Muslims by mapping out their work break schedules over a year (Rodriguez 2018).

Development practitioners and donor countries working on data strategies for public sector reform in countries in which political, religious, or other basic human freedoms are not guaranteed must thus tread carefully in order to guard against setting up data infrastructures that can be used to the detriment of public sector employees.

SYSTEM LEVEL: PRODUCTIVITY, ACCOUNTABILITY, AND INNOVATION

Moving on from the group dimension, this section discusses the most distinctive aspect of data collection on public sector employees: ethical concerns that relate to the duty of public sector employees to serve the public, which can support but also interfere with safeguards designed to protect against unethical data strategies.

Public sector values are commonly defined as the set of qualities of public sector work that make for an ideal public service. Such values typically pertain to productivity (delivering public goods and services in an efficient way), accountability (delivering in a way that is answerable to the public and key stakeholders), and, ever more commonly, innovation (constantly adapting to keep pace with and anticipate societal and economic developments). Each of these qualities can be served by data. The next sections discuss them in more detail.

Other public sector values that are often discussed include equity, transparency, impartiality, and a concern for the common good. As equity and impartiality are supported by mechanisms enforcing accountability and a degree of transparency is required for accountability to be effective, these themes will not be discussed here separately. Similarly, a concern for the common good is difficult to define. As this chapter takes a view built on economic theories that see the common good as the product of welfare-maximizing actions, the next section will discuss the ethics of data collection for the common good together with those aimed at increasing productivity.

Productivity

Public sector workers are meant to maximize productivity in service of the public, in response to their political principals' directives and while remaining accountable to a diverse group of societal stakeholders. In the 21st century, execution is not enough; public sector employees are also expected to do their work as efficiently as possible. They are expected to maximize quality output per tax dollar contributed by each tax-paying citizen. Core to the task of a public sector employee is thus to be productive (for the common social good).

This is not guaranteed.⁶ Public sector workers have a lot of room to diverge from the productive delivery of public service. Public sector workers have specialist knowledge and skills that make it difficult for outsiders to assess the quality and efficiency of their work. A more fast-paced economy and social changes also demand that the public sector be more flexible and responsive, which requires awarding more discretion to public sector workers.

Public sector productivity is difficult to measure because it is a collective effort. There are no market prices readily available for many of the complex goods and services the public sector provides. Efficiency is often a poor marker of success because the services are supplied by the government precisely because there are market failures.

In lieu of rule-based control, oversight in the form of monitoring metrics has become more common. Data and analytics can help overcome the feasibility of, and individual employees' proclivity for, ethical violations. They can ensure that officials are focused on the productivity of public service. Advances in the measurement of productivity have been made, in particular through combining micro- and macro-data, such as process data, project- and task-completion rates, staff and user satisfaction data, performance evaluations, and cost-weighted budget data (Somani 2021).

Data Ethics and Public Sector Productivity

Both accountability and productivity can be promoted by making data on public works projects publicly available and easily understandable. Lauletta et al. (2019) illustrate how a website that geotags public works projects in Colombia can speed up project completion. Several administrations have started using predictive algorithms to allocate inspectors (for example, fire and food safety inspectors) more efficiently. Data can help to promote meritocracy by identifying bottlenecks and resource deficiencies. If data are used to address resource inequalities, they can help public sector workers be productive.

Nehf (2003) uses the term “mischaracterization as harm” to summarize the core of ethical problems related to measuring productivity in the public sector: imperfections in measurement can create misjudgments that are harmful to the dignity of the individuals and groups described as much as they cause more tangible social harm. For instance, when productivity is measured without appropriate awareness of epistemological challenges, it can encourage management to targets. In the context of education, this can lead to grade inflation (De Witte, Geys, and Solondz 2014; Hernández-Julián and Looney 2016). In health care, it has been linked to underinvestment in preventative care (Bevan 2009; Gubb 2009).

Problems with construct validity could unfairly sideline certain groups. For instance, the amounts of sick leave taken and overtime logged are not necessarily good measures of productivity, skill, or work effort. If employees struggle with their health, it infringes on their dignity to equate sick leave with a lack of motivation or commitment to organizational performance. In a similar vein, employees with care responsibilities might be unable or reluctant to work overtime but could nonetheless be stellar performers.

The lack of agreement about what is productive for many job roles in the public sector—or perhaps the agreement that there is no clear definition—means that measurement risks undermining principles of meritocracy. Public services whose productivity is hard to measure might be defunded relative to those whose measurement is easier. Personnel who manage to targets rather than those who create meaningful value for citizens might get promoted. It can also create imbalances in workforce planning. Specialists have been found to be disadvantaged in terms of career progress in the civil service. This seems to be connected to a lack of adequate data on skill matches and to managers’ inability to evaluate what good specialist (as opposed to generalist) performance looks like (Guerin et al. 2021).

An ethical approach to measuring productivity will entail an emphasis on understanding the values that underlie what is considered productive and how power relations shape how problems are defined and acted upon. For example, microdata can also help show where public sector employees might engage in corrupt practices. An ethical approach, however, does not guard against using data on corruption selectively (Nur-tegin and Jakee 2020). Depending on the relative political power of different governing parties and the opposition, data collection efforts might be channeled away from some activities to focus on others. Who has power over data collection efforts and the use of data is thus a question that lies between data capabilities and their possible positive effects on public sector productivity and how ethically public sector personnel are treated.

As discussed in chapter 4 of the *Handbook* effective validation and benchmarking exercises can help to create meaningful and ethically sound measurement of public administration. The chapter argued that a balanced approach to measurement ensures that measurement is meaningful. This chapter argues further that a balanced distribution of power over that measurement and corresponding data will make it more likely that measurement and data are used ethically and justly. Enabling stakeholders to provide checks and balances against data misuse and opportunities for review (see the framework proposed in appendix B) remains key.

Epistemological and practical problems are here to stay. Epistemologically, what is defined as productive is questionable. Questions range from the definitions of core work tasks and what makes an externality to what metric should be used to signal positive impact. Quicker social security claim processing times might signal productivity, or a reduction in maternal mortality at public hospitals might speak to the quality of care, but neither speaks to the dignity with which service users are treated—arguably, an attribute central to public service. Despite improved data quality, which helps to produce efficiencies, value-based decision-making will not become obsolete. Negotiating values is required as much as ever to produce evidence-based and ethically sound policy (Athey 2017).

Accountability

We next turn to the other two key public values that elicit tough ethical challenges for data collection in the public sector: first, accountability, then, innovation. One of the defining characteristics of public sector work is the demand for accountability. Public sector workers are expected to be answerable and liable when they do wrong. The group of people to whom they are answerable is large and diverse. It includes the clients of their services, members of the communities in which their services or policy actions take effect, and organizational stakeholders, such as firms and civil society organizations, whose cooperation might be necessary to arrive at sustainable and equitable policy solutions.

Creating an accountable public administration is a challenging task. The need for delegation and specialization requires that public administrators be provided with discretion. Theory and empirical evidence suggest that political control over the bureaucracy is not a sufficient lever for accountability (Brierley 2020; Gailmard 2002; Keiser and Soss 1998; Meier, Stewart, and England 1991; Raffler 2020; Weingast and Moran 1983; White, Nathan, and Faller 2015).⁷

Democracy in the deep sense—one that goes beyond elections and rests upon an informed and politically active population—requires that policy and its implementation can be scrutinized by the public. The quality of a democracy hinges on the ability of its population to be informed about what state apparatuses do, to voice their opinions about them, and to enforce policy change (Dahl 1998). Bureaucratic accountability also requires transparency and explicability.⁸ As data become more widely available, it becomes easier for experts, civil society groups, and other stakeholders to scrutinize how well the government delivers on providing social goods. Data on the public sector and the civil service thus play an important role in helping to provide accountability.

However, this does not come without challenges. Ethical questions surrounding the use of data for accountability promotion center on the difference between transparency and explicability, concerns surrounding throughput, and the risk of causing unacceptable negative outcomes unintentionally as a result of data management solely focused on external accountability. Such risks require the circumspect creation of data infrastructure.

Data Ethics and Explicability

Explicability, as it is used here, means the quality of being communicated in such a way that most people understand what a thing is or what a process does, its purpose and use. For instance, an algorithm is explicable if the average person could have a heuristic understanding of what it does (for example, rank news items by relevance to the reader based on their previous online behavior, such as clicking on and viewing articles or liking posts). Explicability does not require a person to know the technical details of how this is achieved (for example, the math behind an algorithm's ranking logic). Explicability is thus different from transparency. A transparent algorithm might be open source if everyone could read and check the code that is used to create it, but it is likely not explicable to most people.

A case in point relevant to public services concerns the fact that governments across the globe are increasingly adopting data dashboards that summarize progress on targets for staff and the public. They exemplify how data can provide an easy mechanism that encourages transparency for accountability. For example, dashboards can be used by citizens to monitor the progress of city governments' attempts to improve transport systems by making visible live traffic data, problems, government intervention, and response times (Matheus, Janssen, and Maheshwari 2020). It is, however, important to bear in mind that, like any technological fix, dashboards are no panacea leading to increased accountability (Matheus, Janssen, and Maheshwari 2020). They need to provide information that is actionable for civil society and other stakeholders.

In the context of public service delivery, which increasingly takes a networked or matrix-like form whereby multiple agents within and outside government collaborate to deliver public services, data for accountability need to communicate who responsible parties are and how they can be held to account if they do not deliver on promises. A case in point is the Inter-American Development Bank's MapaInversiones regional initiative, which is "an online platform that allows users to monitor the physical and financial

progress of public investment projects through data visualizations and geo-referenced maps” (Kahn, Baron, and Vieyra 2018, 23). As an evaluation suggested, the project successfully decreased the time taken to start and complete infrastructure projects. Those projects that published their details and progress via the platform fared better than those that did not (Kahn, Baron, and Vieyra 2018, 16).

Another risk that comes with increased transparency in the name of accountability is a decrease in the appetite for innovation. If administrations face a hostile media landscape or political pressure to paint over the challenges they face, an increase in the availability of data on what goes on inside government might stymie innovation. Public sector workers might face a reduction in their freedoms to think creatively and be entrepreneurial. They might be increasingly incentivized to manage to targets and targets only. Citizens would then face an inflexible administration without the capacity and incentive to adapt to changing needs. Thus, we return to the example of data dashboards: their availability must not distract from usability and explicability.

This chapter highlights explicability in particular because there is more demand for transparency regarding the algorithms and unsupervised machine-learning techniques used in public administration (Morley et al. 2020b). Making algorithms and code publicly available increases transparency, but it does not necessarily help citizens understand what they are confronting and how they can call for reform.

Increased data availability that supports accountability must incorporate qualitative aspects, lived experiences, and room for deliberation about what results mean for public service. Open government initiatives (such as the United Nations’ Open Government Data partnership and the World Bank’s Anti-corruption, Openness, Transparency, and Political Economy Global Solutions Group) and unilateral government efforts (such as the Open Knowledge Foundation) to make government data more accessible, explicable, and usable for a diverse group of stakeholders are good cases in point for how accountability can be at the center of data collection efforts.

Throughput

Throughput describes behavior that “makes something happen.” This contrasts with plans, values, or ideals that relate to action but are not the actions that create the announced change. For instance, having a team of legal staff who regularly check contracts and account details to verify that employees with comparable skills and experience levels receive the same pay is the difference between having equal pay policies and providing throughput on them.

Where data concern the attitudes, opinions, and experiences of public sector staff, using these data for accountability promotion requires throughput. Surveying service providers (public sector workers in this case) or service users can make organizations appear considerate and proactive. However, if momentum and resources are lacking to enact change based on survey results, what is intended as an accountability mechanism can soon amount to window dressing. This is problematic in terms of monetary value—the time taken from staff and clients to answer surveys goes wasted—and in terms of trust in institutions. If accountability mechanisms are not used as promised, they can backfire. They can create mistrust, disengagement, and cynicism where they intended to foster trust, engagement, and optimism.

Unintended Consequences

For public accountability, a government should know how much it spends on its workforce and who gets what. Many jurisdictions make the incomes of public sector employees public. The disclosure of salaries has propelled efforts to close gender disparities in pay (Marcal 2017). It might subsequently seem innocuous to track data on pay and incentives.

However, organizations typically know more than how much an employee earns. Many employers, particularly public sector employers, offer benefits such as health insurance and pension schemes, sometimes even bursaries for the educational and medical needs of employees’ families. From the types of pension and insurance arrangements chosen by an employee, an employer can easily see what types of health issues an employee might face and how many dependents profit from the employee’s benefits. This can create unintended breaches of privacy in the name of accountability.

For instance, while it is admirable that employers subsidize additional insurance schemes designed to cover extra costs associated with mental illness or cancer treatment, typically not covered by basic insurance packages, this also means that employers hold data on the mental and physical health of staff. Holding such data increases the risk associated with data breaches and data misuse. On top of the broad types of information on health an employer might glean from insurance choices, organizations that opt to provide their workforce with fitness trackers face an even more granular level of data and associated risk.

It can be argued that public sector employees can be subjected to greater levels of scrutiny, with more of their personal data accessible to public view, because they inhabit positions of power and the trade-off of privacy for power is just. We see such a trade-off with people who are considered to be public personas—politicians and celebrities—and it might not amount to overstepping to extend this category to include public sector employees. What is considered an infringement on privacy is, however, deeply embedded in social context and norms (Regan 2018). The dilemma created in this situation is that two sets of social norms are at odds: the belief that a person's address, name, income, and work activities should not be freely available for consultation on the internet versus the belief that this is justified when the individual belongs to the category of "public persona."

Navigating these dilemmas sustainably requires that, when individuals join social groups in which different norms on data privacy are applied, they do so knowingly and are provided with the opportunity to advocate to change these norms. In practice, this can mean giving public sector staff unions a voice in what data are made available about their members and what guarantees of accountability they can offer the public in place of such data.

Innovation

Gaining an improved understanding of how public administrations perform is central to promoting evidence-based innovation. Promises of revolutionizing government via increased data collection, analysis, and insights abound. The World Bank released a report in early 2021 focused solely on the role of data in promoting growth and international development (World Bank 2021b).

In Indonesia, for instance, survey, census, administrative, and satellite data are being combined to help to plan infrastructure and improve urban sustainability (World Bank Data Team 2019). A report published in 2021 suggests that data on postal mail could be used to help stem drug smuggling and counterfeit trade (Shivakumar 2021). Furthermore, HR data were successfully used by Colonnelli, Prem, and Teso (2020) to capture patterns of politicized hiring and firing in the Brazilian public service. Election, media, and administrative and accountability data, such as scorecards, highlight where governments deliver on their promises of responsiveness and service delivery and where they lag behind (Brierley 2020; Erlich et al. 2021; Ingrams, Piotrowski, and Berliner 2020). The *Handbook* includes examples of how data on public administrators can help to create happier and more effective workplaces (chapters 1 and 2), flag malpractice (chapter 7), better assess the quality of government processes (chapter 13), and combat systemic institutional biases (chapter 20).²

Welfare Consequentialism

In the context of data collection for public sector innovation, with a particular focus on data on the inner workings of government, ethical problems mainly surround two sets of questions. First are those concerned with welfare consequentialism. As outlined above, most public sector discourse stresses how data can be used for welfare gains through the alleviation of poverty, energy savings, or similar large-scale changes. Using such a logic, however, raises questions about *whose* benefit is targeted, who decides what is valuable, how valuable certain ends are, and whether, for a process to be considered ethical, evaluation should focus on actual or intended and expected consequences. Such concerns apply to the measurement of public administration as well as its outputs. It is commonplace in many countries for public sector employers to garner data on their recruits' police records, previous addresses (going back many years), social media accounts and public postings, credit scores, and other data as part of routine background checks in hiring

processes (see Erlam [2008] for a review of European and US law and practices regarding background checks). Is the end of all this data collection and consolidation a more effective recruitment process? Or are data sets routinely collected that impinge on officials' privacy but do not speak to their effectiveness?

Garnering insights for innovation will often require accumulating and linking data that have previously not been accumulated, linked, or analyzed in the same manner. Risks associated with this include the potential to breach the rights to privacy and confidentiality of individuals or employee groups. This can occur through poorly designed data collection, usage, and storage protocols, data breaches, or data vending strategies. The most defensible position to take for an evaluation of the morality of such data strategies in the public sector is one that defines what is morally right according to whether its *intended* as well as its *actual* consequences create good outcomes. In other words, systems should be designed with a view to preventing potential harm, and action plans should be available to mitigate or stop harmful actions when they are underway.

A case in point that highlights the need for a systems perspective on the intended and unintended consequences of using data for public sector innovation is the UK National Health Service (NHS). The NHS decided to sell individual and linked health data to third-party companies in order to use analytical skills outside government to help plan, forecast, and reform health care services. However, there have been doubts about whether the sale of data can be conducted in a way that prevents third parties from misusing these data—for instance, for marketing purposes (Rahman 2021; Rapp and Newman 2021; Rector 2021). It is also questionable whether, despite anonymizing details of the data (such as names and addresses), individuals and groups are protected from reidentification via triangulation with other data sources. Another example is provided by the departments of motor vehicles (DMVs) in the US states of Texas and Colorado, which sold service user data to commercial companies (New 2020; Whalen 2019). While the sale happened under the terms and conditions agreed to by service users, some of the sales were used for identity theft and document forgery by third parties who had legally acquired the data (Lieber 2020).

As these examples illustrate, while data protocols for the innovative use of public sector data will involve working with third sector parties, mission drift as much as the intentional misuse of data needs to be considered when designing data protocols. The examples also illustrate that the existence of rules and, in the case of the GDPR (which was still in force when the NHS data system was set up), the threat of high legal fines are not enough to guarantee that data systems generate usage patterns that can be widely accepted as ethical.

The most appealing solution to this problem in democratic contexts would be to involve service users and interested parties in working groups that are used to adapt systems. Apart from the hurdles inherent to all participatory approaches, such a solution faces the challenge that innovative uses of data more often than not involve new technologies, pregnant with new jargon to decipher. Creating data approaches for public sector innovation thus requires that time and resources be set aside to make the process explicable to those affected by it. This is no simple task because governments still face skills gaps in cutting-edge areas of information technology. In many instances, governments will need to rely on external expertise to develop and maintain the skills of their staff to implement data solutions that are ethically sound and secure.¹⁰

Innovation and the Core Mandates of the Public Sector

There is a danger that a greater push for data availability for innovation's sake will conflict with other demands on the public sector. Innovation in the short run can look unproductive: there are high costs and little immediate return.

Innovation and accountability can be at odds. For example, data on innovations related to defense, energy, and telecommunications infrastructures cannot be made readily available without weighing risks to national security. It is also unclear what rate of failure is acceptable for a public sector organization. Cost-benefit arguments that downweight the ethical risks associated with increased data collection and tracking efforts based on promises associated with the potential for public sector innovation should include such limitations in order not to overstate potential benefits.

This second set of concerns relates to how the imperative to innovate interacts with other core mandates of the public sector, such as enforcing rules, protecting fundamental rights, and stepping in where there are

market failures. Jordan (2014) postulates that the moral imperative for public servants to act within their assigned responsibilities is greater than the imperative for innovation, so that even when they willingly block innovation but stay within their assigned mandate, their guilt is limited. Such a claim becomes highly problematic as soon as we move outside a context where every group in society has equal access to enforcing accountability. Even in countries where electoral democracy flourishes, state institutions can be systematically biased against certain groups. Mere rule-following can then create significant welfare costs.

Over the last decade, the use of data-driven policing has provided many examples of how a genuine conviction to render public services more efficient and automatized can negatively affect swaths of society. Profiling algorithms have been shown, in some cases, to hide systematic discrimination deep inside their processing structures. In these situations, it is arguably more ethically acceptable for public servants to go out of their way to improve algorithms and change policing rather than block changes to the status quo because they do not fit their narrow job description.

When data collection for innovation concerns data on public servants themselves—colleagues, superiors, and supervisees—resistance to innovation seems equally misplaced. For instance, if public servants willfully hinder the collection of data on gender intended to promote gender-equitable hiring and promotion practices, they harm fellow citizens (their colleagues) by denying them access to services that other parts of government granted them, and they potentially harm society by obstructing reforms of the civil service that could help to improve society as a whole.

It should be noted that what is considered a transgression and what is considered part of a public servant's sphere of discretion is subject to debate—a debate that can be highly politicized. It would hence be most appropriate to revise Jordan's (2014) idea of a moral imperative to rule-following over innovation to a moral imperative to guard public welfare over rule-following, including directives related to innovation.

The age-old problem is that in the absence of rules and moral absolutes, the state apparatus needs to rely on the moral compasses of individuals and the norms reinforced by organizational culture to navigate toward the ethical delivery of public service. Appendix B therefore tries to provide a framework for public servants to evaluate new data approaches.

CONCLUSION

While guides on research ethics abound, there is little guidance available for how ethical data collection, analysis, and innovation with data on public sector workers should take place. This chapter has attempted to provide such guidance via a heuristic that sees ethical challenges as falling broadly within three dimensions: an individual dimension, which comprises demands for dignity and privacy; a group dimension, which relates to voice and dissent; and a public-facing dimension, which ensures that data enable public administrators to deliver on public sector values. Each of these dimensions affords similar but also different prerogatives. The individual dimension calls for dignity and personal privacy, the group dimension relates to how data create and diminish voice, and the public-facing dimension concerns how data relate to public sector values. There is a wide range of values and considerations that can be framed as “public value.” This chapter has focused on three that are central to many discussions of public administration's effectiveness, efficiency, and measurement: productivity, accountability, and innovation. The section on productivity highlighted how important it is to choose metrics well and understand their strengths and biases (see more on this in sections 2 and 3 of the *Handbook*). The discussion of accountability presented the tensions in using data to increase the accountability of the public service by emphasizing explicability over mere transparency. The discussion of the tensions inherent to using data to create and measure innovation, as well as the delicate balance between accountability and innovation, showed that dialogue and regular reviews of the data strategies adopted by governments and researchers to measure public sector work and support innovation must be baked into the process to guard against delivering more to one value than another.

To make things more practical and support the reflective approach to measuring public sector work and employee effort, in appendix B we offer a framework that practitioners can use to build and review measurement strategies.

NOTES

This chapter is indebted to extensive comments provided by Josh Cows (University of Oxford) and Johannes Himmelreich (Syracuse University) on dimensions of big data ethics and to Dan Rogger (World Bank) and Christian Schuster (University College London) on public administration and bureaucracy.

1. In the public sector, the right to individual dignity is curtailed by the public's right to dignity. This tension is explored in more detail further along in this chapter, in the section on the public dimension of data collection ethics.
2. For instance, in Qatar, 54 percent of the population is employed by the government. While employing half of the population is rare, employing 25–30 percent of the population is the norm in most middle-income and Western European countries. Data are from the International Labour Organization ILOSTAT “ILO modelled estimates database” (accessed January 2021), ilostat.ilo.org/data, via World Bank Open Data, “Employment in industry (% of total employment) (modeled ILO estimate),” World Bank, Washington, DC, <https://data.worldbank.org/indicator/SL.IND.EMPL.ZS?>
3. Action research is a form of research that uses mostly qualitative methodologies but can also involve the creation of survey questionnaires and the scales used to quantify responses. It is defined by an emphasis on making research participatory—involving the subjects of research and data collection actively in all stages of the research process, from defining the research question and the parameters of data collection to the use of data and results (for a more in-depth explanation, see, for example, Brydon-Miller, Greenwood, and Maguire 2003).
4. Note that *citizens* here is meant also to encompass persons without formal claims to citizenship who are government service users or fall within the jurisdiction of the government.
5. The theory being that these children perform worse typically as they struggle to fit in and potentially face discrimination. Minority teachers are hypothesized to be more cognizant of these problems and to create an atmosphere that is more inclusive and nurturing for minority pupils.
6. This dilemma, dubbed the “principal-agent problem,” has been widely discussed and is still researched in a variety of ways in economics, political science, and public administration.
7. For example, in some contexts, it has been demonstrated that political control over the bureaucracy increases instances of corruption and other malpractice (Brierley 2020), while in others, it can improve practices (Raffler 2020).
8. Lavertu (2016) sees an active role for public administration scholars in contextualizing, revising, and creating metrics to provide a more holistic assessment of public sector performance, in order to prevent misguided intervention by political and citizen principals.
9. As alluded to in earlier sections, a lot of rhetoric on public sector reform invokes an imperative to innovate. The underlying assumption is usually consequentialist: innovation will create benefits; therefore, it is good. Public sector innovation's merit can also be framed in terms of the collective virtue of creativity, an endeavor that is intrinsically worth pursuing, and extending it via a hedonistic logic, for the positive life experiences it can create.
10. Some governments have already developed schemes to encourage skills transfer into the government from outside for innovative practices. For example, Latvia runs a “shadow an entrepreneur” program, as part of which civil servants learn from private and third sector entrepreneurs about new developments in tech (OECD OPSI 2020).

REFERENCES

- Alesina, Alberto, and Eliana La Ferrara. 2005. “Ethnic Diversity and Economic Performance.” *Journal of Economic Literature* 43 (3): 762–800. <https://www.jstor.org/stable/4129475>.
- Athey, Susan. 2017. “Beyond Prediction: Using Big Data for Policy Problems.” *Science* 355 (6324): 483–85. <https://doi.org/10.1126/science.aal4321>.
- Baer, Markus, and Michael Frese. 2003. “Innovation Is Not Enough: Climates for Initiative and Psychological Safety, Process Innovations, and Firm Performance.” *Journal of Organizational Behavior* 24 (1): 45–68. <https://doi.org/10.1002/job.179>.
- Bala, Sumathi. 2021. “Rise in Online Payments Spurs Questions over Cybersecurity and Privacy.” *CNBC*, July 1, 2021. <https://www.cnbc.com/2021/07/01/newdigital-payments-spur-questions-over-consumer-privacy-security-.html>.

- Basu, Tanya. 2020. "Vote Count Livestreams Are Here to Stay: Sit Back and Enjoy the Show as You Watch Election Results Get Tallied." *MIT Technology Review*, November 4, 2020. <https://www.technologyreview.com/2020/11/04/1011648/livestream-vote-counts-are-here-to-stay/>.
- BBC News. 2019. "Wikileaks: Document Dumps That Shook the World." April 12, 2019. <https://www.bbc.com/news/technology-47907890>.
- BBC News. 2021. "Conservative Conference: Get Off Your Pelotons and Back to Work, Says Oliver Dowden." October 5, 2021. <https://www.bbc.com/news/uk-politics-58804607>.
- Beioley, Kate. 2022a. "Brexit Dividend' Rule Change Prompts Fears over Data Flow with EU." *Financial Times*, February 28, 2022. <https://www.ft.com/content/8da85688-76f0-4a97-9e58-f701f3488d78>.
- Beioley, Kate. 2022b. "Metaverse vs Employment Law: The Reality of the Virtual Workplace." *Financial Times*, February 20, 2022. <https://www.ft.com/content/9463ed05-c847-425d-9051-482bd3a1e4b1>.
- Bevan, Gwyn. 2009. "Have Targets Done More Harm Than Good in the English NHS? No." *BMJ* 338: a3129. <https://doi.org/10.1136/bmj.a3129>.
- 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>.
- Boot, Eric R. 2020. "The Feasibility of a Public Interest Defense for Whistle-Blowing." *Law and Philosophy* 39 (1): 1–34. <https://doi.org/10.1007/s10982-019-09359-1>.
- Brierley, Sarah. 2020. "Unprincipled Principals: Co-opted Bureaucrats and Corruption in Ghana." *American Journal of Political Science* 64 (2): 209–22. <https://doi.org/10.1111/ajps.12495>.
- Brydon-Miller, Mary, Davydd Greenwood, and Patricia Maguire. 2003. "Why Action Research?" *Action Research* 1 (1): 9–28. <https://doi.org/10.1177/14767503030011002>.
- Buell, Ryan W., Tami Kim, and Chia-Jung Tsay. 2017. "Creating Reciprocal Value through Operational Transparency." *Management Science* 63 (6): 1673–95. <https://doi.org/10.1287/mnsc.2015.2411>.
- Cabinet Office. 2022. *Civil Service Sickness Absence, 2021: Report*. London: UK Government. <https://www.gov.uk/government/publications/civil-service-sickness-absence-2021/civil-service-sickness-absence-2021-report>.
- Card, David, Thomas Lemieux, and W. Craig Riddell. 2020. "Unions and Wage Inequality: The Roles of Gender, Skill and Public Sector Employment." *Canadian Journal of Economics/Revue canadienne d'économique* 53 (1): 140–73. <https://doi.org/10.1111/caje.12432>.
- Carpenter, David. 2017. "The Quest for Generic Ethics Principles in Social Science Research." In *Finding Common Ground: Consensus in Research Ethics across the Social Sciences*, edited by Ron Iphofen, 3–17. Advances in Research Ethics and Integrity 1. Bingley, UK: Emerald Publishing. <https://doi.org/10.1108/S2398-601820170000001001>.
- CBS and the Associated Press. 2015. "Millions More Government Fingerprints Deemed Stolen." *CBS News*, September 23, 2015. <https://www.cbsnews.com/news/millions-more-government-fingerprints-deemed-stolen/>.
- Clark, Pilita. 2021. "Employee Data Can Be Used for Good, but Treat It with Care." *Financial Times*, July 24, 2021. <https://www.ft.com/content/b9cf429e-5529-47cf-88e0-7122aaed0240>.
- Cofone, Ignacio N. 2020. *The Right to Be Forgotten: A Canadian and Comparative Perspective*. New York: Routledge.
- Colonnelli, Emanuele, Mounu Prem, and Edoardo Teso. 2020. "Patronage and Selection in Public Sector Organizations." *American Economic Review* 110 (10) (October): 3071–99. <https://doi.org/10.3886/E118165V1>.
- Cooper, Christopher A. 2020a. "Impartiality and Public Sector Employees' Online Political Activity: Evidence from Three Australian Elections." *Acta Politica* 57: 210–34. <https://doi.org/10.1057/s41269-020-00181-5>.
- Cooper, Christopher A. 2020b. "Public Servants, Anonymity, and Political Activity Online: Bureaucratic Neutrality in Peril?" *International Review of Administrative Sciences* 86 (3): 496–512. <https://doi.org/10.1177/0020852318780452>.
- Dahl, Robert Alan. 1998. *On Democracy*. New Haven, CT: Yale University Press.
- De Witte, Kristof, Benny Geys, and Catharina Solondz. 2014. "Public Expenditures, Educational Outcomes and Grade Inflation: Theory and Evidence from a Policy Intervention in the Netherlands." *Economics of Education Review* 40: 152–66. <https://doi.org/10.1016/j.econedurev.2014.02.003>.
- Düwell, Marcus, Jens Braarvig, Roger Brownsword, and Dietmar Mieth, eds. 2014. *The Cambridge Handbook of Human Dignity: Interdisciplinary Perspectives*. Cambridge: Cambridge University Press.
- Erlam, N. Alexander. 2008. "Understanding International Background Checks in the Age of Privacy and Data Security Concerns." *International In-House Counsel Journal* 1 (3): 489–98. <https://www.iicj.net/subscribersonly/08april/iicjapr4-data-pricay-nalexandererlam-vericalscreen.pdf>.
- Erlich, Aaron, Daniel Berliner, Brian Palmer-Rubin, and Benjamin E. Bagozzi. 2021. "Media Attention and Bureaucratic Responsiveness." *Journal of Public Administration Research and Theory* 31 (4) (February): 687–703. <https://doi.org/10.1093/jopart/muab001>.
- Ford, Michael R. 2022. "A Little Representation Goes a Long Way: Minority Teacher Representation and District Performance in a Highly Homogenous Context." *Public Organization Review* 22: 691–705. <https://doi.org/10.1007/s11115-021-00535-3>.

- Fung, Brian. 2019. "Lawmakers Want to Ban 'Dark Patterns,' the Web Designs Tech Companies Use to Manipulate You." *Washington Post*, April 9, 2019. <https://www.washingtonpost.com/technology/2019/04/09/policymakers-are-sounding-alarm-dark-patterns-manipulative-web-design-trick-youve-never-heard/>.
- Gailmard, Sean. 2002. "Expertise, Subversion, and Bureaucratic Discretion." *Journal of Law, Economics, and Organization* 18 (2): 536–55. <https://www.jstor.org/stable/3555054>.
- Grantham, Kathleen. 2020. *Mapping Gender Data Gaps: An SDG Era Update*. Washington, DC: Data2x.
- Grisson, Jason A., Samantha L. Viano, and Jennifer L. Selin. 2016. "Understanding Employee Turnover in the Public Sector: Insights from Research on Teacher Mobility." *Public Administration Review* 76 (2): 241–51. <https://doi.org/10.1111/puar.12435>.
- Gubb, James. 2009. "Have Targets Done More Harm Than Good in the English NHS? Yes." *BMJ* 338: a3130. <https://doi.org/10.1136/bmj.a3130>.
- Guerin, Benoit, Alex Thomas, Rhys Clyne, and Suhasini Vira. 2021. *Finding the Right Skills for the Civil Service*. London: Institute for Government. <https://www.instituteforgovernment.org.uk/sites/default/files/publications/civil-service-skills.pdf>.
- Heath, Joseph, and Wayne Norman. 2004. "Stakeholder Theory, Corporate Governance and Public Management: What Can the History of State-Run Enterprises Teach Us in the Post-Enron Era?" *Journal of Business Ethics* 53 (3): 247–65. <https://doi.org/10.1023/B:BUSI.0000039418.75103.ed>.
- Hernández-Julián, Rey, and Adam Looney. 2016. "Measuring Inflation in Grades: An Application of Price Indexing to Undergraduate Grades." *Economics of Education Review* 55: 220–32. <https://doi.org/10.1016/j.econedurev.2016.11.001>.
- Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Cambridge, MA: Harvard University Press.
- Hjort, Jonas. 2014. "Ethnic Divisions and Production in Firms." *The Quarterly Journal of Economics* 129 (4): 1899–946. <https://doi.org/10.1093/qje/qju028>.
- Hunter, Tatum. 2021. "Here Are All the Ways Your Boss Can Legally Monitor You." *Washington Post*, September 24, 2021. <https://www.washingtonpost.com/technology/2021/08/20/work-from-home-computer-monitoring/>.
- Ingrams, Alex, Suzanne Piotrowski, and Daniel Berliner. 2020. "Learning from Our Mistakes: Public Management Reform and the Hope of Open Government." *Perspectives on Public Management and Governance* 3 (4) (December): 257–72. <https://doi.org/10.1093/ppmgov/gvaa001>.
- Israel, Mark. 2015. *Research Ethics and Integrity for Social Scientists: Beyond Regulatory Compliance*. 2nd ed. London: Sage. <https://doi.org/10.4135/9781473910096>.
- Izzo, Zachary, Mary Anne Smart, Kamalika Chaudhuri, and James Zou. 2021. "Approximate Data Deletion from Machine Learning Models." In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, edited by Arindam Banerjee and Kenji Fukumizu, 2008–16. Proceedings of Machine Learning Research 130. <http://proceedings.mlr.press/v130/izzo21a.html>.
- Jordan, Sara R. 2014. "The Innovation Imperative: An Analysis of the Ethics of the Imperative to Innovate in Public Sector Service Delivery." *Public Management Review* 16 (1): 67–89. <https://doi.org/10.1080/14719037.2013.790274>.
- Kahn, Theodore, Alejandro Baron, and J. Cruz Vieyra. 2018. "Digital Technologies for Transparency in Public Investment: New Tools to Empower Citizens and Governments." Discussion Paper IDB-DP-634, Inter-American Development Bank, Washington, DC. <http://dx.doi.org/10.18235/0001418>.
- Keiser, Lael R., and Joe Soss. 1998. "With Good Cause: Bureaucratic Discretion and the Politics of Child Support Enforcement." *American Journal of Political Science* 42 (4): 1133–56. <https://doi.org/10.2307/2991852>.
- Kenny, Kate. 2019. *Whistleblowing: Toward a New Theory*. Cambridge, MA: Harvard University Press.
- Kingsley, J. Donald. [1944] 2003. "Representative Bureaucracy." In *Representative Bureaucracy: Classic Readings and Continuing Controversies*, edited by Julie Dolan and David H. Rosenbloom, 12–18. New York: Routledge.
- Lauletta, Maximiliano, Martín Antonio Rossi, Juan Cruz Vieyra, and Diego Arisi. 2019. "Monitoring Public Investment: The Impact of MapaRegalias in Colombia." Working Paper IDB-WP-1059, Inter-American Development Bank, Washington, DC. <https://ideas.repec.org/p/idb/brikps/9967.html>.
- Lavertu, Stéphane. 2016. "We All Need Help: 'Big Data' and the Mismeasure of Public Administration." *Public Administration Review* 76 (6): 864–72. <https://doi.org/10.1111/puar.12436>.
- Lieber, Dave. 2020. "After 27 Million Driver's License Records Are Stolen, Texans Get Angry with the Seller: The Government." *Dallas Morning News*, November 26, 2020. <https://www.dallasnews.com/news/watchdog/2020/11/26/after-27-million-drive-rs-license-records-are-stolen-texans-get-angry-with-the-seller-government/>.
- Lynn, Matthews. 2022. "If We Sacked Working-from-Home Civil Servants, Would Anyone Notice?" *Telegraph*, May 12, 2022. <https://www.telegraph.co.uk/news/2022/05/12/sacked-working-from-home-civil-servants-would-anyone-notice/>.
- Macdonald, Sharon. 2017. "Embedded Ethics and Research Integrity: A Response to 'The Quest for Generic Ethics Principles in Social Science Research' by David Carpenter." In *Finding Common Ground: Consensus in Research Ethics across the Social Sciences*, edited by Ron Iphofen, 29–35. Advances in Research Ethics and Integrity 1. Bingley, UK: Emerald Publishing. <https://doi.org/10.1108/S2398-60182017000001003>.

- Madrigal, Alexis C. 2012. "Reading the Privacy Policies You Encounter in a Year Would Take 76 Work Days." *Atlantic*, March 1, 2012. <https://www.theatlantic.com/technology/archive/2012/03/reading-the-privacy-policies-you-encounter-in-a-year-would-take-76-work-days/253851/>.
- Marcal, Katrin. 2017. "Sweden Shows That Pay Transparency Works." *Financial Times*, July 27, 2017. <https://www.ft.com/content/2a9274be-72aa-11e7-93ff-99f383b09ff9>.
- Matheus, Ricardo, Marijn Janssen, and Devender Maheshwari. 2020. "Data Science Empowering the Public: Data-Driven Dashboards for Transparent and Accountable Decision-Making in Smart Cities." *Government Information Quarterly* 37 (3): 101284. <https://doi.org/10.1016/j.giq.2018.01.006>.
- Mehr, Hila. 2017. *Artificial Intelligence for Citizen Services and Government*. Ash Center for Democratic Governance and Innovation, Harvard Kennedy School, Cambridge, MA. <https://ash.harvard.edu/files/ash/files/artificialintelligenceforcitizen-services.pdf>.
- Meier, Kenneth J. 2019. "Theoretical Frontiers in Representative Bureaucracy: New Directions for Research." *Perspectives on Public Management and Governance* 2 (1) (March): 39–56. <https://doi.org/10.1093/ppmgov/gvy004>.
- Meier, Kenneth J., Joseph Stewart Jr., and Robert E. England. 1991. "The Politics of Bureaucratic Discretion: Educational Access as an Urban Service." *American Journal of Political Science* 35 (1): 155–77. <https://doi.org/10.2307/2111442>.
- Miller, Gary J., and Andrew B. Whitford. 2016. *Above Politics*. Cambridge, UK: Cambridge University Press.
- MIT Technology Review and Oracle. 2016. *The Rise of Data Capital*. Cambridge, MA: MIT Technology Review Custom in Partnership with Oracle. <http://files.technologyreview.com/whitepapers/MITOracle+Report-TheRiseofDataCapital.pdf>.
- Mohan, Bhavya, Ryan W. Buell, and Leslie K. John. 2020. "Lifting the Veil: The Benefits of Cost Transparency." *Marketing Science* 39 (6): 1105–21. <https://doi.org/10.1287/mksc.2019.1200>.
- Moore, Mark H. 1995. *Creating Public Value: Strategic Management in Government*. Cambridge, MA: Harvard University Press.
- Morley, Jessica, Josh Cowls, Mariarosaria Taddeo, and Luciano Floridi. 2020a. "Ethical Guidelines for COVID-19 Tracing Apps." *Nature* 582: 29–31. <https://doi.org/10.1038/d41586-020-01578-0>.
- Morley, Jessica, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2020b. "From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices." *Science and Engineering Ethics* 26 (4): 2141–68. <https://doi.org/10.1007/s11948-019-00165-5>.
- Nehf, James P. 2003. "Recognizing the Societal Value in Information Privacy." *Washington Law Review* 78 (1): 1–91. <https://digitalcommons.law.uw.edu/wlr/vol78/iss1/2>.
- Nembhard, Ingrid M., and Amy C. Edmondson. 2006. "Making It Safe: The Effects of Leader Inclusiveness and Professional Status on Psychological Safety and Improvement Efforts in Health Care Teams." *Journal of Organizational Behavior* 27 (7): 941–66. <https://doi.org/10.1002/job.413>.
- New, Brian. 2020. "I-Team: Texas DMV Made USD 3M Last Year Selling Drivers' Personal Information." *CBS Dallas-Fort Worth*, February 20, 2020. <https://dfw.cbslocal.com/2020/02/20/texas-dmv-selling-drivers-personal-information/>.
- Nur-tegin, Kanybek, and Keith Jakee. 2020. "Does Corruption Grease or Sand the Wheels of Development? New Results Based on Disaggregated Data." *The Quarterly Review of Economics and Finance* 75: 19–30. <https://doi.org/10.1016/j.qref.2019.02.001>.
- OECD (Organisation for Economic Co-operation and Development). 2001. *Citizens as Partners: OECD Handbook on Information, Consultation and Public Participation in Policy-Making*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264195578-en>.
- OECD (Organisation for Economic Co-operation and Development). 2019. *The Path to Becoming a Data-Driven Public Sector*. OECD Digital Government Studies. Paris: OECD Publishing. <https://www.oecd-ilibrary.org/governance/the-path-to-becoming-a-data-driven-public-sector059814a7-en>.
- OECD (Organisation for Economic Co-operation and Development). 2022. *Recommendation of the Council Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*. OECD/LEGAL/0188. Paris: OECD Publishing.
- OECD OPSI (Organisation for Economic Co-operation and Development Observatory of Public Sector Innovation). 2020. *Embracing Innovation in Government: Global Trends 2020: Upskilling and Investing in People*. Paris: OECD Publishing. <https://trends.oecdopsi.org/trend-reports/upskilling-and-investing-in-people/>.
- Office for National Statistics. 2022. "Sickness Absence in the UK Labour Market: 2021." UK Government, April 29, 2022. <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/articles/sicknessabsenceinthelabourmarket/2021>.
- Office for Science. 2020. *Evidence and Scenarios for Global Data Systems: The Future of Citizen Data Systems*. London: UK Government. <https://www.gov.uk/government/publications/the-future-of-citizen-data-systems>.
- Open Data Institute. 2021. *Data Ethics: How to Be More Trustworthy with Data*. <https://theodi.org/service/consultancy/data-ethics/>.
- Pamuk, Humeyra, and Simon Lewis. 2021. "U.S. State Department Names Former Ambassador Gina Abercrombie-Winstanley as First Chief Diversity Officer." *Reuters*, April 12, 2021. <https://www.reuters.com/article/us-usa-diversity-diplomacy-idUSKBN2BZ2AD>.

- Panagiotopoulos, Panos, Bram Klievink, and Antonio Cordella. 2019. "Public Value Creation in Digital Government." *Government Information Quarterly* 36 (4): 101421. <https://doi.org/10.1016/j.giq.2019.101421>.
- Preston, Larry M. 1987. "Freedom and Bureaucracy." *American Journal of Political Science* 31 (4): 773–95. <https://doi.org/10.2307/2111224>.
- Raffler, Pia. 2020. "Does Political Oversight of the Bureaucracy Increase Accountability? Field Experimental Evidence from a Dominant Party Regime." *American Political Science Review* 116 (4): 1443–59. <https://doi.org/10.1017/S0003055422000181>.
- Rahman, Grace. 2021. "What's Happening with Your NHS Data?" *Full Fact*, June 8, 2021. <https://fullfact.org/health/nhs-data/>.
- Rapp, Ben, and Sara Newman. 2021. "NHS Digital's GP Data-Scraping Plan Must Be Publicised and Delayed." *Computer Weekly*, June 7, 2021. <https://www.computerweekly.com/opinion/NHS-Digitals-GP-data-scraping-plan-must-be-publicised-and-delayed>.
- Rasul, Imran, and Daniel Rogger. 2015. "The Impact of Ethnic Diversity in Bureaucracies: Evidence from the Nigerian Civil Service." *American Economic Review* 105 (5): 457–61. <https://doi.org/10.1257/aer.p20151003>.
- Rector, Alan. 2021. "The NHS Data Grab: Why We Should Be Concerned about Plans for GPs' Records." *Guardian*, June 6, 2021. <https://www.theguardian.com/society/2021/jun/06/the-nhs-data-grab-why-we-should-be-concerned-about-plans-for-gps-records>.
- Regan, Priscilla. 2018. "Legislating Privacy: Technology, Social Values, and Public Policy." In *The Handbook of Privacy Studies: An Interdisciplinary Introduction*, edited by Bart van der Sloot and Aviva de Groot, 57–62. Amsterdam: Amsterdam University Press. <https://doi.org/10.1515/9789048540136-003>.
- Regan, Priscilla M. 2000. *Legislating Privacy: Technology, Social Values, and Public Policy*. Chapel Hill, NC: University of North Carolina Press.
- Resnik, David B. 2020. "Ethical Issues in Research Involving Employees and Students." In *The Oxford Handbook of Research Ethics* [online edition], edited by Ana S. Iltis and Douglas MacKay. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190947750.013.40>.
- Reuters. 2022. "Nvidia Says Employee, Company Information Leaked Online after Cyber Attack." March 1, 2022. <https://www.reuters.com/technology/nvidia-says-employee-company-information-leaked-online-after-cyber-attack-2022-03-01/>.
- Rhodes, Rod A. W., and John Wanna. 2007. "The Limits to Public Value, or Rescuing Responsible Government from the Platonic Guardians." *Australian Journal of Public Administration* 66 (4): 406–21. <https://doi.org/10.1111/j.1467-8500.2007.00553.x>.
- Riivari, Elina, and Anna-Maija L"ams"a. 2014. "Does It Pay to Be Ethical? Examining the Relationship between Organisations' Ethical Culture and Innovativeness." *Journal of Business Ethics* 124 (1): 1–17. <https://doi.org/10.1007/s10551-013-1859-z>.
- Rodriguez, N. 2018. "Expanding the Evidence Base in Criminology and Criminal Justice: Barriers and Opportunities to Bridging Research and Practice." *Justice Evaluation Journal* 1 (1): 1–14. <https://doi.org/10.1080/24751979.2018.1477525>.
- Schmutte, Ian M., and Lars Vilhuber. 2020. "Balancing Privacy and Data Usability: An Overview of Disclosure Avoidance Methods." In *Handbook on Using Administrative Data for Research and Evidence-Based Policy*. New Haven, CT: Yale University Press. <https://admindatahandbook.mit.edu/book/v1.0/discavoid.html>.
- Schroeder, Doris, and Abol-Hassan Bani-Sadr. 2017. "The Quest for Dignity." In *Dignity in the 21st Century: Middle East and West*, 1–8. Cham, Switzerland: Springer.
- Selin Davis, Lisa. 2021. "The Right to Be Forgotten: Should Teens' Social Media Posts Disappear as They Age?" *Washington Post*, June 14, 2021. <https://www.washingtonpost.com/lifestyle/2021/06/14/should-what-children-post-online-come-back-haunt-them-later-life/>.
- Shivakumar, Sujai. 2021. "Better Data on International Mail Packages Could Reduce Illegal Drugs and Counterfeit." Center for Data Innovation, May 17, 2021. <https://datainnovation.org/2021/05/better-data-on-international-mail-packages-could-reduce-illegal-drugs-and-counterfeits/>.
- Solove, Daniel J. 2008. *Understanding Privacy*. Cambridge, MA: Harvard University Press.
- Somani, Ravi. 2021. *Public-Sector Productivity (Part One): Why Is It Important and How Can We Measure It?* Washington, DC: World Bank. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/913321612847439794/public-sector-productivity-part-one-why-is-it-important-and-how-can-we-measure-it>.
- Sommer, David Marco, Liwei Song, Sameer Wagh, and Prateek Mittal. 2020. "Towards Probabilistic Verification of Machine Unlearning." arXiv:2003.04247v2 [cs.CR]. <https://doi.org/10.48550/arXiv.2003.04247>.
- Stone, Deborah A. 2002. *Policy Paradox: The Art of Political Decision Making*. New York: W. W. Norton.
- Suderman, Alan. 2021. "DC Police Victim of Massive Data Leak by Ransomware Gang." Associated Press, May 13, 2021. <https://apnews.com/article/police-technology-government-and-politics-1aedfc42a8dc2b004ef610d0b57edb9>.
- Taylor, Linnet, Luciano Floridi, and Bart van der Sloot. 2016. *Group Privacy: New Challenges of Data Technologies*. Cham, Switzerland: Springer.
- UK Government. 2015. "The Civil Service Code." March 16, 2015. <https://www.gov.uk/government/publications/civil-service-code/the-civil-service-code>.

- UN Data Revolution Group (United Nations Secretary-General's Independent Expert Advisory Group (IEAG) on a Data Revolution for Sustainable Development). 2014. *A World That Counts: Mobilising the Data Revolution for Sustainable Development*. UN Data Revolution Group. <https://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf>.
- Vidgen, Richard, Giles Hindle, and Ian Randolph. 2020. "Exploring the Ethical Implications of Business Analytics with a Business Ethics Canvas." *European Journal of Operational Research* 281 (3): 491–501. <https://doi.org/10.1016/j.ejor.2019.04.036>.
- Viswanath, Shilpa. 2020. "Public Servants in Modern India: Who Are They?" In *The Palgrave Handbook of the Public Servant*, edited by Helen Sullivan, Helen Dickinson, and Hayley Henderson, 1–16. Cham, Switzerland: Springer. <https://doi.org/10.1007/978-3-030-03008-796-1>.
- Wang, Ray. 2012. "What a Big-Data Business Model Looks Like." *Harvard Business Review*, December 7, 2012. <https://hbr.org/2012/12/what-a-big-data-business-model>.
- Weingast, Barry R., and Mark J. Moran. 1983. "Bureaucratic Discretion or Congressional Control? Regulatory Policymaking by the Federal Trade Commission." *Journal of Political Economy* 91 (5): 765–800. <https://www.jstor.org/stable/1837369>.
- Whalen, Andrew. 2019. "DMVs across the Country Selling Your Driver's License Data for as Little as a Penny, Making Them Millions." *Newsweek*, September 6, 2019. <https://www.newsweek.com/dmv-drivers-license-data-database-integrity-department-motor-vehicles-1458141>.
- White, Ariel R., Noah L. Nathan, and Julie K. Faller. 2015. "What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials." *American Political Science Review* 109 (1): 129–42. <https://doi.org/10.1017/S0003055414000562>.
- Williams, Sophie, and Frances Mao. 2022. "Winter Olympics: Athletes Advised to Use Burner Phones in Beijing." *BBC News*, January 18, 2022. <https://www.bbc.co.uk/news/world-asia-china-60034013>.
- World Bank. 2021a. *Europe and Central Asia Economic Update, Spring 2021: Data, Digitalization, and Governance*. Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/35273>.
- World Bank. 2021b. *World Development Report 2021: Data for Better Lives*. Washington, DC: World Bank. <https://www.worldbank.org/en/publication/wdr2021>.
- World Bank Data Team. 2019. "7 Data Innovation Projects Win Funding to Tackle Local Challenges." *Data Blog. World Bank Blogs*, January 19, 2019. <https://blogs.worldbank.org/opendata/7-data-innovation-projects-win-funding-tackle-local-challenges>.
- Wright, R. George. 1987. "Speech on Matters of Public Interest and Concern." *DePaul Law Review* 37 (1): 27–52. <https://via.library.depaul.edu/law-review/vol37/iss1/3>.

CHAPTER 7

Measuring and Encouraging Performance Information Use in Government

Donald Moynihan

SUMMARY

Public sector organizations can leverage government analytics to improve the quality of their services and internal functioning. But the existence of government analytics does not ensure its use. This chapter discusses how governments can measure the use of government analytics—or, in other words, conduct analytics on government analytics. The chapter assesses, in particular, one important component of government analytics: *performance management systems*. It does so by, first, contrasting different types of *performance information use*, including political and purposeful uses. The chapter next assesses different approaches to measure performance information use, including those based on surveys, administrative data, and qualitative inquiry. Finally, the chapter illustrates how the measurement of performance information use can be built over time and drawn upon to improve government analytics, using the example of the US Government Accountability Office (GAO). The chapter's findings underscore the importance of the robust measurement of government analytics use to enable governments to make the most of government analytics.

ANALYTICS IN PRACTICE

- The existence of government analytics systems does not automatically translate to the use of government analytics. Measurement does not equate to the use of measures for management action. As a result, policy makers should pay close attention not only to whether government analytics is used but also to exactly how it is used.
- The use of government analytics should be purposeful and strategic, but it can often fall prey to passive, political, and perverse uses. Ideally, performance information should be used by government agencies to purposefully plan for achieving their goals, either through internal restructuring or decision-making.

Donald Moynihan is the McCourt Chair at the McCourt School of Public Policy, Georgetown University.

However, when the design of *performance management systems* generates distortionary incentives, these same systems can lead public officials to perversely manipulate data or use performance information for political gain. This puts a premium on measuring *performance information use* to understand such (ab)use.

- Government analytics use can be measured using administrative data, surveys, qualitative data, and decision-based inferences. Administrative data can track how frequently performance information (for example, on a dashboard) is accessed; surveys of public servants can inquire about the extent of the use of performance information and about different types of use; qualitative data (for example, through focus groups) can provide rich detail on how (elite) decision-making is shaped by performance information; and decision-based inferences can showcase how informing public servants or managers about information alters their decision-making.
- The effective measurement of government analytics use benefits from longer time horizons that enable learning from previous experiences, as well as reforms to strengthen government analytics. In the United States, the stability of the core elements of a performance management system has meant that measurement can track progress in performance information use over time in order to help reformers understand what works and what needs fixing and gradually incorporate those lessons into practice.

INTRODUCTION

Government analytics enables governments to diagnose and improve public management across the public administration production function (see chapter 1). Yet the measurement of the internal machinery of government does not necessarily translate into the actual use of government analytics to improve public administration. This puts a premium on encouraging the use of government analytics, as a number of chapters in *The Government Analytics Handbook* discuss (see, for example, chapters 25 and 26). It also, however, underscores the importance of measuring whether—and how—government analytics is used by public sector organizations to improve their functioning. In other words, it underscores the importance of doing analytics on the use of government analytics.

This chapter discusses how this can be done, focusing on the case of *performance management systems*. Governments rely on performance management systems, sets of coordinated processes to measure, validate, disseminate, and make use of performance data within the government. Early government performance systems focused on performance measurement and have gradually transitioned, with varying degrees of success, to performance management (Moynihan and Beazley 2016). Performance management systems have often emphasized routines of data collection, including requirements to set strategic goals and short-term targets, and have measured some combination of outputs and outcomes. But the collection of performance information does not, by itself, ensure its use (Kroll 2015).

How can governments monitor progress in the use of performance measurements? This chapter provides guidance in this effort. First, it provides a conceptual grounding, discussing different types of *performance information use*, including political and purposeful uses. Subsequently, it moves to measurement: different ways of measuring performance information use, including engagement metrics based on administrative data, survey-based instruments, and qualitative inquiry. Finally, it illustrates, using the case of the US Government Accountability Office (GAO), how to construct effective governmentwide measures of performance information use and underscores their utility to improve performance management systems.

THE (AB)USES OF PERFORMANCE INFORMATION: A TYPOLOGY

Performance management systems often struggle to demonstrate their own effectiveness. As a result, there is remarkably thin evidence on the causal effects of performance management systems on government performance

itself (Gerrish 2016). One important intermediate outcome of performance management systems, however, can be measured, and this is performance information use. By *performance information use*, I refer to the extent to which data are considered by and inform the decisions and behavior of public officials (Kroll 2015).

Performance information can be construed as an intermediate measure of the effectiveness of performance management systems. Performance measures cannot generate improved performance by themselves. To have an effect, they have to alter the judgment and decision-making of a human being. In other words, data must be effectively used for management. If data are left unused, it is impossible for them to improve public sector outcomes through changes in the behavior of civil servants.

Types of Performance Information Use

Conceptually, it is useful to distinguish between four distinct types of performance information use (Kroll 2015; Moynihan 2009). Box 7.1 lays out four primary types of performance information use, along with the scholarly literature informing their definitions.

BOX 7.1 Types of Performance Information Use

Purposeful: The central hope of performance management reformers is that public employees use data to improve program performance. Such improvements can come via goal-based learning that gives rise to efficiency improvements, better targeting of resources, and more informed strategic decisions, or by tying indicators to rewards or sanctions in contract arrangements.

Passive: Performance management reforms may result in passive reactions, in which officials do the minimum required to comply with requirements to create and disseminate information but do not actually use this information (Radin 2006). This approach is also more likely in hierarchical relationships, where actors often lack strong incentives to use data but are not penalized for not using them. Cynicism based on failed performance management reforms in the past increases the likelihood of a passive response because the current reform will be perceived as temporary. A passive response is also more likely if elected officials, stakeholders, and agency leaders demonstrate little real interest in implementing performance management tools. Where results-based reforms have a permanent statutory basis, it becomes more difficult for public servants to assume they can wait them out.

Political: Performance management reforms are grounded in a demand for public agencies to present evidence that they are performing. As a result, public employees may see performance information as a means to define their efforts and success. Performance data thereby become a means of advocacy in a political environment (Moynihan 2008). In many cases, employees have some degree of influence in selecting and measuring the performance goals by which they are judged. They are likely to select, disseminate, and interpret information that portrays them in a favorable light.

Perverse: In some cases, the pressures to maximize measured performance may be so great that agents will improve these measures in ways that are in conflict with the underlying goals of a program. Agents may game program indicators through a variety of tactics, including making up data, selecting easy-to-serve clients while neglecting more difficult ones (*cream-skimming*), focusing on measured goals at the expense of important unmeasured goals or values (*goal displacement*), changing performance goals to limit comparison across time, or manipulating measures (Courty and Marschke 2004). Gaming becomes more likely when strong financial, career, or reputational incentives are attached to performance indicators, as well as when measures only partially capture the underlying goal.

Source: Adapted from Moynihan (2009).

The purposeful use of performance information anchors the use of performance indicators in goal orientation, improving resource allocation and strategic decisions. As such, the purposeful use of performance indicators can improve the efficiency of public administration. There is, therefore, strong interest in creating an environment that pushes public officials from passive to purposeful performance information use, as well as in monitoring for and minimizing perverse forms of use.

Public servants are, in general, the primary users of performance data. Members of the public might support the idea of a performance-driven government in the abstract but cannot be expected to be active users of public sector performance data. The only settings in which the public may be more interested in performance data are those where there is market-based choice in the provision of services or where services have high personal salience to the public. Some examples are the choice of schools or medical providers. Elected officials may express interest in performance data and seek them out in specific domains but are generally reluctant to let them alter their ideological preferences or other factors that shape their decisions (Christensen and Moynihan 2020; Moynihan 2008).

Beyond making data available and attractive, designers of performance management systems ultimately have little control over whether members of the public or policy makers use data, but they have much more control over the environment of public servants and thus greater opportunity to shape that environment to ensure performance data are used. Public servants also enjoy extensive program-level knowledge that makes them more likely to understand how to make sense of detailed performance information that would be difficult for nonspecialist actors to understand. Thus, I focus on public servants' use of performance data.

HOW IS PERFORMANCE INFORMATION USE MEASURED?

Performance information use is a cognitive behavior, a form of judgment and decision-making that is often difficult to observe. Methodological approaches to measuring the use of performance systems require design choices, each associated with trade-offs. I briefly consider these approaches and trade-offs before concluding that surveys are the most practical option for governments seeking to track the use of performance management systems, although administrative data and qualitative approaches can provide important complements.

Engagement-Based Inferences on Performance Information Use, Using Administrative Data

The first—and most obvious—approach to measuring the extent of performance information use is to observe directly whether public officials engage with existing performance management systems. These engagement metrics can include, for instance, whether officials download performance data or use them in public discussions or reports.

This approach has the advantage of not relying on subjective reports and can be used in a nonintrusive way. At the same time, governments must be willing to invest in tracking real-time data on engagement by public officials with performance information. This often requires having a centralized website where information is made available. For example, Andersen and Moynihan (2016) observe that school principals are more likely to download available performance data under certain experimental conditions, including when they have been delegated more authority in hiring decisions and offered comparative information that they find more useful.

Such an approach may be especially useful to observe how public servants engage with the performance management system. There is a set of indicators available for doing so: how many times performance information has been downloaded, how much time public officials spend engaging with the data in, for example, a dashboard, and how many times performance information has been shared on social media, such as Twitter or organizational forums.

Engagement-based inference requires data quality checks and may become less reliable if employees know their behavior is being observed. Additionally, engagement measures can be biased toward frequent

but unimpactful use. After all, such measures provide information on the frequency of use but not on how performance information is used, for instance, in decision-making. A single instance of performance information use by a senior authority in an organization may result in a *low-frequency* measure in a dashboard tracking time spent on a performance dashboard. Yet this instance may be more impactful in terms of triggering organizationwide management changes than dozens of employees' consulting the dashboard.

Survey-Based Inferences on Performance Information Use

A second popular approach to evaluate the use of performance data is through public servant surveys. To discuss survey-based inferences in practice, I highlight the example of the United States. In the US federal government, the GAO is responsible for measuring performance information use. A central part of the GAO's strategy is a series of periodic surveys of public employees about performance management. These have taken place in 1997, 2000, 2003, 2007, 2013, 2017, and 2020 (US GAO 1997, 2001, 2004, 2008, 2013, 2017, 2021). The result has been a series of insightful reports that track how well the performance management system is being implemented and offer suggestions for improvement (for example, US GAO 2005, 2008, 2013, 2014).

Over time, the GAO has used a consistent set of indicators to track different types of performance information use. This has allowed the GAO to assess variation in performance information use across agencies. The results of the surveys show variation consistent with logical sources (discussed below). Survey-based inferences are discussed below in the context of the US performance management system. Box 7.2 presents a few examples of the precise wording of the GAO survey question on the use of the performance management system, as well as potential indicators as they correspond to box 7.1 on the types of performance information use.

BOX 7.2 US Government Accountability Office Survey Measures of Performance Information Use

For those program(s)/operation(s)/project(s) that you are involved with, to what extent, if at all, do you use the information obtained from performance measurement when participating in the following activities? Responses range from "to no extent" (0) to "to a very great extent" (4).

Passive performance information use

- Refining program performance measures
- Setting new or revising existing performance goals

Purposeful performance information use

- Developing program strategy
- Setting program priorities
- Allocating resources
- Identifying program problems to be addressed
- Taking corrective action to solve program problems
- Adopting new program approaches or changing work processes
- Coordinating program efforts with other internal or external organizations
- Identifying and sharing effective program approaches with others
- Developing and managing contracts
- Setting individual job expectations for the government employees the respondent manages or supervises
- Rewarding government employees that the respondent manages or supervises

Source: Adapted from US GAO (2021).

Qualitative Inferences on Performance Information Use

Beyond quantitative measures of performance information use, qualitative assessments can provide contextual knowledge on how performance information is being used in government. Qualitative assessments include focus groups and structured interviews in which public servants are invited to share their experiences with performance information use. During these qualitative assessments, interviewers enable public servants and managers to share their own stories regarding performance information use.

Qualitative assessments thus provide a complementary perspective to measures of how public servants use performance information. Qualitative assessments provide a holistic assessment of how performance information use is embedded within organizational routines, public servants' perceptions, and their performance of everyday tasks. While information on organizational functioning may be collected through public servant surveys as well, interviews and focus groups provide an insider's perspective and insights into elite decision-making based on performance information within organizations, and they embed them within a narrative.

This narrative—for example, “managers do not use performance information”—may be explored using other sources of data. A narrative may also validate whether performance information is having its intended effect (purposeful use) or is being used perversely or politically within the government. For this reason—much in the spirit of Bridges and Woolcock in chapter 4 of this *Handbook*—qualitative inquiry can complement survey and administrative data to enable analysts to gain a more holistic understanding of the use of government analytics.

Decision-Based Inferences on Performance Information Use

A final approach to measuring performance information use is to evaluate whether providing decision-makers with performance information changes observed behavior. For example, a decision-maker might be asked to make a judgment about a program or organization or a decision about budgets, management, or a contract. Researchers then observe how much this decision varies depending upon the presence or absence of performance data, variation in the content of the performance data, changes in the environment in which the data have been presented, and individual traits.

As might be apparent, the controlled nature of such studies tends to rely on hypothetical scenarios, even if researchers sometimes try to increase their generalizability by using actual policy makers or public managers as respondents and actual government analytics data. This artificiality is one constraint of this mode of study and one reason it is often not practical in trying to assess general patterns of performance information use among civil servants (though survey experiments could certainly be inserted into such surveys).

Despite these potential limitations, policy experiments introducing decision-makers to informational treatments on performance have generated a significant body of work. For example, Holm (2018) measures how performance indicators affect management decisions by analyzing how public school principals define school priorities after being informed of performance results. Indeed, he finds that school principals prioritize goals with lower performance and devote their efforts to improvement in these lagging areas.

It becomes more difficult to evaluate whether performance information affects decision-making at a governmentwide level. One reason is that the bountiful array of performance data that exists makes it difficult to isolate the influence of one piece of information. Some researchers have looked at budget recommendations by government officials and budget decisions by legislators to see how these are correlated with whether summative performance data were positive or negative.

For example, there is clear evidence that program-level performance scores issued by the George W. Bush administration in the United States affected the budget recommendations it made to Congress, with less clear evidence that Congress followed those recommendations (Moynihan 2013). In parliamentary systems, where the executive and legislature work together, there is a greater likelihood that performance-informed executive judgments match legislative decisions, according to some studies (Sohn, Han, and Bae 2022). Decision-based inference is an important tool for understanding whether performance information changes

behavior. However, it should be a complement to rather than a substitute for administrative data and survey-based or qualitative inquiries, which provide a relatively more holistic understanding of performance information use.

Potential Biases from Performance Information Use

It is important to note that, beyond the typology of deliberate uses of performance information by decision-makers (for example, purposeful or perverse use), decision-makers may be subject to biases when consulting performance information (James et al. 2020). For example, there is ample evidence for the power of negativity bias: policy makers and members of the public pay more attention to negative rather than positive performance data (Nielsen and Moynihan 2017; Olsen 2015).

As a result, providing decision-makers with performance information and measuring their deliberate use is not enough: biases, ideological or political, may affect how decision-makers use this information (Baekgaard and Serritzlew 2016; Christensen et al. 2018; James and Van Ryzin 2017; Nielsen and Moynihan 2017). There is ample evidence that decision-makers select performance information that fits with their ideological beliefs and discount data that do not (James and Van Ryzin 2017). Policy makers are more resistant to efforts to reduce such biases than the general public, reflecting their status as more skilled and dedicated motivated reasoners (Christensen and Moynihan 2020).

CASE APPLICATION: THE GAO'S ROLE IN THE MODERNIZATION ACT

The US federal government currently uses a performance management system designed under the Government Performance and Results Act Modernization Act (GPRAMA), enacted in 2010. The Office of Management and Budget (OMB), part of the Executive Office of the President, is charged with leading the implementation of the system, while the GAO, part of the legislative branch, is charged with monitoring it. Several lessons about how to systematically measure performance information use can be drawn from this experience—which is, arguably, the world's most sustained effort to measure performance information use over time across government.

Lesson One: An independent agency evaluating progress through a consistent set of measurement tools enhances credibility and knowledge accumulation.

The GAO has issued reports evaluating progress in performance management and the use of performance information to Congress and made them publicly available. Some have been qualitative case analyses of performance reforms in particular agencies or focus groups of senior managers across agencies, but the primary source of information is the survey data collected from employees across the federal government over time.¹ The credibility of the GAO has helped ensure a high response rate and thus enhanced the credibility of the resulting information on performance information use in the US federal government.

Lesson Two: Evaluations of progress over time in performance information use benefit from stability in performance information.

The stability of the performance management system has allowed for comparisons across time. The surveys of federal government employees and managers have asked the same core questions over time. The GAO has also added questions that reflect changes in the performance management system while keeping the basic performance information use measures intact. This has allowed for analysis of the relationship between

aspects of the performance management system that change with a standard set of items over time. The data thus show, for instance, that while early evaluations of the system were relatively critical, they have become more positive over time, reflecting a process of gradual learning (Moynihan and Kroll 2016).

Lesson Three: Data-sharing agreements with external researchers can provide additional insights into measures of performance information use.

GAO reports have provided valuable insights into the progress of the performance management system. But they have also been able to generate substantial additional research and insight by soliciting suggestions from outside researchers about what questions to ask and by sharing the data when they were collected. These data have enabled researchers to generate a series of analyses. Such analyses are not constrained by the format of government reports and can ask different types of questions, combine survey with nonsurvey data, and use more advanced statistical methods. Federal government officials from both the OMB and the GAO have been responsive to this research, incorporating insights into budget and planning documents.

The GPRAMA also has lessons for encouraging performance information use more generally, complementing other chapters in the *Handbook* on this topic (for example, chapter 26). Because performance information was repeatedly measured in the US case, there is a stronger (survey) evidence base on the factors that encourage performance information use. This reinforces lesson three about sharing performance information use data with researchers. For instance, Moynihan and Kroll (2016; see also Kroll and Moynihan 2021) find that exposure to high-quality organizational routines—such as agency leaders’ identifying high-priority, agency-specific or cross-agency goals or the leadership team’s reviewing progress toward key goals every quarter, based on well-run, data-driven reviews—matters to performance information use.

Moynihan and Lavertu (2012) find that leaders who engage seriously with performance management systems encourage followers to do the same and that employees who have more discretion to make decisions are more likely to use performance data. Kroll and Moynihan (2015) also find that exposure to performance management training is associated with higher use, while Kroll and Moynihan (2018) find that managers who use program evaluations are more likely to use performance data. This has become particularly salient since the Obama administration, which built bridges between program evaluation research and performance managers in government, partly by institutionalizing performance staff, such as performance improvement officers, who view their job more broadly than reporting performance data.

As these examples underscore, analytics of repeated measurement of performance information use can help governments understand this use, the factors driving it, and how to improve performance management systems over time to further increase their use. The analytics of performance information use data can thus ensure governments make the most of their performance management systems.

CONCLUSION

The existence of government analytics does not ensure its effective use. This chapter shows that government analytics data can be used in a range of ways. Data can be used purposefully for management improvements, as intended by the designers of government analytics systems. Data can also be misused politically, however, or perverted, distorting organizational incentives.

Government analytics should thus not only measure whether and how frequently data are being used but also for what purpose. Much like for government analytics generally, a range of data sources are available for such analytics, including administrative data (for example, tracking users on a dashboard), survey data (for example, surveys of public employees and managers), and qualitative inquiry, among others.

The case of the United States—arguably the world’s most sustained effort to measure performance information use over time across government—underscores that such analytics is particularly helpful with

certain design features. For instance, the US system benefits from a stable set of performance information use measures and the regular collection of data assessing progress but also from a community of practice interested in using this longer-run data on government analytics to understand which parts of performance information (or government analytics) are working and what needs fixing. This community of career officials within the executive branch, who oversee performance issues, GAO officials in Congress, who complete their analyses, and external researchers, who share their findings, has then seen the gradual incorporation of lessons into practice.

The analytics of government analytics use can help improve and institutionalize how the government uses performance data to generate better outcomes. Ultimately, government analytics can only change the way governments operate if public officials meaningfully engage with, analyze, and operationalize analytical insights from data. This chapter thus underscores the importance of robust measurement of the use of government analytics to enable governments to make the most of the tools at their disposal.

NOTE

1. Rather than trying to survey all employees, the GAO engaged in random sampling of mid- and senior-level federal managers. In 1997 and 2000, it used mail surveys, but in 2003, it transitioned to email-based surveys linked to a web questionnaire (with follow-up phone efforts for nonrespondents).

REFERENCES

- Andersen, S. C., and D. P. Moynihan. 2016. "Bureaucratic Investments in Expertise: Evidence from a Randomized Controlled Field Trial." *Journal of Politics* 78 (4): 1032–44.
- Baekgaard, M., and S. Serritzlew. 2016. "Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension." *Public Administration Review* 76 (1): 73–82.
- Christensen, J., C. M. Dahmann, A. H. Mathiasen, D. P. Moynihan, and N. B. G. Petersen. 2018. "How Do Elected Officials Evaluate Performance? Goal Preferences, Governance Preferences, and the Process of Goal Reprioritization." *Journal of Public Administration Research and Theory* 28 (2): 197–211.
- Christensen, J., and D. P. Moynihan. 2020. "Motivated Reasoning and Policy Information: Politicians Are More Resistant to Debiasing Interventions than the General Public." *Behavioural Public Policy*, FirstView, 1–22. <https://doi.org/10.1017/bpp.2020.50>.
- Courty, P., and G. Marschke. 2004. "An Empirical Investigation of Gaming Responses to Performance Incentives." *Journal of Labor Economics* 22 (1): 23–56.
- Gerrish, E. 2016. "The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis." *Public Administration Review* 76 (1): 48–66.
- Holm, J. M. 2018. "Successful Problem Solvers? Managerial Performance Information Use to Improve Low Organizational Performance." *Journal of Public Administration Research and Theory* 28 (3): 303–20.
- James, O., A. L. Olsen, D. P. Moynihan, and G. G. Van Ryzin. 2020. *Behavioral Public Performance: How People Make Sense of Government Metrics*. Cambridge, UK: Cambridge University Press.
- James, O., and G. G. Van Ryzin. 2017. "Motivated Reasoning about Public Performance: An Experimental Study of How Citizens Judge the Affordable Care Act." *Journal of Public Administration Research and Theory* 27 (1): 197–209.
- Kroll, A. 2015. "Drivers of Performance Information Use: Systematic Literature Review and Directions for Future Research." *Public Performance and Management Review* 38 (3): 459–86.
- Kroll, A., and D. P. Moynihan. 2015. "Does Training Matter? Evidence from Performance Management Reforms." *Public Administration Review* 75 (3): 411–20.
- Kroll, A., and D. P. Moynihan. 2018. "The Design and Practice of Integrating Evidence: Connecting Performance Management with Program Evaluation." *Public Administration Review* 78 (2): 183–94.
- Kroll, A., and D. P. Moynihan. 2021. "Tools of Control? Comparing Congressional and Presidential Performance Management Reforms." *Public Administration Review* 81 (4): 599–609.

- Moynihan, D. P. 2008. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC: Georgetown University Press.
- Moynihan, D. P. 2009. "Through a Glass, Darkly: Understanding the Effects of Performance Regimes." *Public Performance and Management Review* 32 (4): 592–603.
- Moynihan, D. P. 2013. "Advancing the Empirical Study of Performance Management: What We Learned from the Program Assessment Rating Tool." *American Review of Public Administration* 43 (5): 499–517.
- Moynihan, D. P., and I. Beazley. 2016. *Toward Next-Generation Performance Budgeting: Lessons from the Experiences of Seven Reforming Countries*. Directions in Development: Public Sector Governance. Washington, DC: World Bank.
- Moynihan, D. P., and A. Kroll. 2016. "Performance Management Routines That Work? An Early Assessment of the GPRA Modernization Act." *Public Administration Review* 76 (2): 314–23.
- Moynihan, D. P., and S. Lavertu. 2012. "Does Involvement in Performance Reforms Encourage Performance Information Use? Evaluating GPRA and PART." *Public Administration Review* 72 (4): 592–602.
- Nielsen, P. A., and D. P. Moynihan. 2017. "How Do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy." *Journal of Public Administration Research and Theory* 27 (2): 269–83.
- Olsen, A. L. 2015. "Citizen (Dis)satisfaction: An Experimental Equivalence Framing Study." *Public Administration Review* 75 (3): 469–78.
- Radin, B. A. 2006. *Challenging the Performance Movement*. Washington, DC: Georgetown University Press.
- Sohn, H., D. Han, and K. B. Bae. 2022. "The Effect of Performance Information on Actual Budgeting Allocation: The Role of Policy Type." *Public Performance and Management Review* 45 (5): 1112–32. <https://doi.org/10.1080/15309576.2022.2062014>.
- US GAO (US General Accounting Office). 1997. *The Government Performance and Results Act: 1997 Governmentwide Implementation Will Be Uneven*. Report to Congressional Committees, GAO-97-109. Washington, DC: US GAO.
- US GAO (US General Accounting Office). 2001. *Managing for Results: Federal Managers' Views on Key Management Issues Vary Widely across Agencies*. Report to the Chairman, Subcommittee on Oversight of Government Management, Restructuring, and the District of Columbia, Committee on Governmental Affairs, US Senate, GAO-01-592. Washington, DC: US GAO.
- US GAO (US General Accounting Office). 2004. *Results-Oriented Government: GPRA Has Established a Solid Foundation for Achieving Greater Results*. Report to Congressional Requesters, GAO-04-38. Washington, DC: US GAO.
- US GAO (US Government Accountability Office). 2005. *Performance Budgeting: PART Focuses Attention on Program Performance, but More Can Be Done to Engage Congress*. Report to the Chairman, Subcommittee on Government Management, Finance, and Accountability, Committee on Government Reform, House of Representatives, GAO-06-28. Washington, DC: US GAO.
- US GAO (US Government Accountability Office). 2008. *Government Performance: Lessons Learned for the Next Administration on Using Performance Information to Improve Results*. Statement of Bernice Steinhardt, Director Strategic Issues. Testimony before the Subcommittee on Federal Financial Management, Federal Services, and International Security, Committee on Homeland Security and Governmental Affairs, US Senate, GAO-08-1026T. Washington, DC: US GAO.
- US GAO (US Government Accountability Office). 2013. *Managing for Results: Executive Branch Should More Fully Implement the GPRA Modernization Act to Address Pressing Governance Challenges*. Report to Congressional Committees, GAO-13-518. Washington, DC: US GAO.
- US GAO (US Government Accountability Office). 2014. *Managing for Results: Agencies' Trends in the Use of Performance Information to Make Decisions*. Report to Congressional Addressees, GAO-14-747. Washington, DC: US GAO.
- US GAO (US Government Accountability Office). 2017. *Managing for Results: Further Progress Made in Implementing the GPRA Modernization Act, but Additional Actions Needed to Address Pressing Governance Challenges*. Report to Congressional Committees, GAO-17-775. Washington, DC: US GAO.
- US GAO (US Government Accountability Office). 2021. *Evidence-Based Policymaking: Survey Results Suggest Increased Use of Performance Information across the Federal Government*. Report to Congressional Committees, GAO-22-103910. Washington, DC: US GAO.

CHAPTER 8

Understanding Corruption through Government Analytics

James Anderson, David S. Bernstein, Galileu Kim, Francesca Recanatini, and Christian Schuster

SUMMARY

Corruption is a multidimensional phenomenon that affects governments and citizens across the world. Recent advances in data collection and analytics have generated new possibilities for both detecting and measuring corruption. This chapter illustrates how the public sector production function introduced in *The Government Analytics Handbook* helps holistically conceptualize where corruption can occur in public administration. It then outlines how corruption can be detected in its multiple dimensions using the microdata approaches outlined in the remaining chapters of this book. Along the production function, corruption can be detected with input data (for example, personnel or budget data), data about processes (for example, survey data on management practices), and output and outcome data (for example, public service delivery data). Using corruption as a thematic focus, the chapter thus showcases how the approaches presented in the *Handbook* can be combined and leveraged to holistically diagnose a key issue in public administration. The chapter complements this methodical discussion with a broader consideration of how political-economic constraints affect policy reforms to reduce corruption.

ANALYTICS IN PRACTICE

- Corruption is a multidimensional phenomenon, affecting public administration across its parts. Corruption can affect public administration in personnel and payroll, through patronage appointments of public servants, and through the payment of ghost workers. Corruption can disrupt service delivery if public servants demand bribes in exchange for access to public services, such as health care and education. Since corruption affects public administration in various ways, a holistic assessment of corruption requires multiple data sources and methodologies.

Galileu Kim is a research analyst for the World Bank's Development Impact Evaluation (DIME) Department. James Anderson is a lead governance specialist for the World Bank's Governance Global Practice (GGP). David S. Bernstein is a lead public sector specialist for the GGP. Francesca Recanatini is a lead economist for the GGP. Christian Schuster is a professor at University College London.

- Recent advances in data collection and analytics have generated new possibilities to detect and measure corruption. The use of data and indicators on corruption is a long-standing tradition in international development. Previous efforts primarily relied on expert assessments and national-level indicators, which are susceptible to subjective bias and lack a foundation in microdata. More recently, there has been growth in administrative microdata, such as procurement, payroll, and other data sources, like surveys of public servants. These rich data environments create novel possibilities to engage in more granular analytics for measuring corruption.
- The public sector production function structures holistic conceptualization and measurement of corruption in public administration. This chapter showcases how conceptualizing public administration as a production function with distinct parts enables corruption to be unpacked into its different dimensions, such as procurement and service delivery. In procurement, for instance, corruption can take the form of the capture of bidding processes by companies or the bribery of public officials. To measure this type of corruption, procurement data can be analyzed to create red flags on bids with a single bidder.
- Since corruption cuts across the public administration production function, the integration of the different data sources and methodologies presented in other chapters in *The Government Analytics Handbook* enables a comprehensive assessment of how corruption operates. For instance, corruption by public servants can stem from different causes. Public servants who engage in corruption might be dissatisfied with their wages or pressured by their managers. Measuring this complex environment of agents and organizational dynamics requires multiple data sources. For instance, managerial pressure can be measured through public servant surveys, while payroll data can provide a sense of pay equity.
- Measurement of corruption can guide and complement public sector reforms, but it is not a substitute for the implementation of challenging policies to reduce corruption. Measuring where corruption occurs can guide public sector reform by detecting areas of vulnerability—for example, ghost workers—and informing reforms—for example, improving quality control and payroll enforcement. Measurement cannot substitute for the important step of implementing challenging policy reforms that will likely be resisted by agents who benefit from the status quo. Reformers should be cognizant of the political-economic environment, which may deter reforms from taking place.
- The multidimensional analytical approach presented in this chapter can be leveraged for other key topics in public administration. While the thematic focus of this chapter is corruption, we emphasize that other important topics in public administration can also benefit from the application of analytical approaches based on multiple data sources. For example, performance management could leverage survey data on public sector motivation and management practices, as well as administrative data on performance indicators such as delays in processing business licenses. Using multiple data sources and analytical approaches enables a more holistic understanding of how public administration maps onto these key issues.

INTRODUCTION

Corruption in public administration has many faces (Campos and Pradhan 2007).¹ To name just two, corruption affects how public services work through nepotism and patronage—the use of political and personal connections for professional gain (Colonnelli, Prem, and Teso 2020; World Bank 2021). It can also shape how state resources are allocated, diverting funds from public education or health for private gain (for example, Ferraz, Finan, and Moreira 2012). This puts a premium on better understanding, measuring, and fighting against corruption. Throughout this chapter, we follow a common definition of corruption as “the use of public office for private gain” (Jain 2001; Rose-Ackerman 1978).

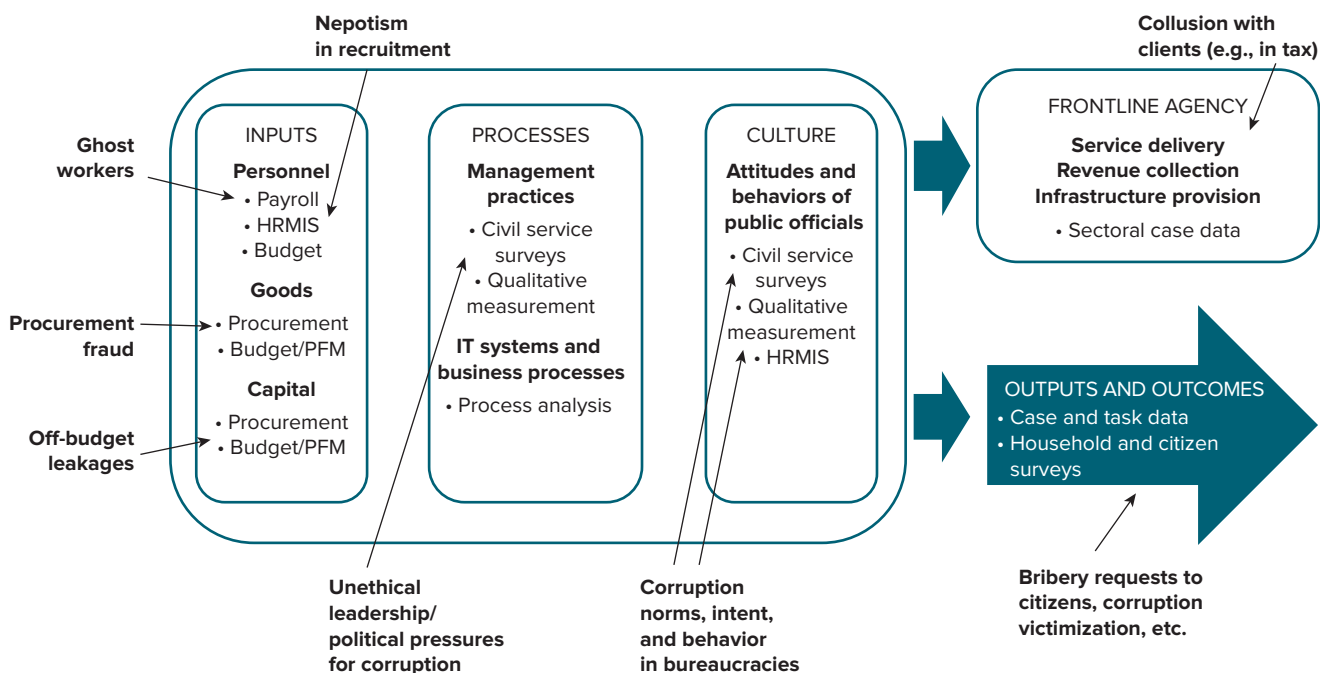
The use of government analytics—and data more broadly—to fight corruption is a long-standing tradition in the international development community (see, for example, Kaufmann, Pradhan, and Ryterman 1998).

Initiatives such as the World Bank’s World Governance Indicators (WGI) and Transparency International’s Corruption Perception Index (CPI) have sought to aggregate a set of indicators into a single measure to help draw attention to issues of governance and the control of corruption. These indicators often rely on expert surveys or qualitative indicators that provide national-level indicators on corruption. More recently, there has been growth in microdata analyzing corruption. For example, surveys of public servants provide staff-level perspectives on how corruption operates within countries and across sectors and government agencies (Recanatini 2011). The growing availability of microdata—administrative and survey-based—provides novel opportunities to increase and refine analytical approaches to understanding and reducing corruption.²

In this chapter, we demonstrate how to leverage the microdata sources and methodologies described in *The Government Analytics Handbook* to measure corruption. We do so through the public administration production function (figure 8.1). Corruption can be measured along the production function with input data on personnel—for example, human resources management information systems (HRMIS)—data about processes in the management of public servants—for example, surveys of public servants—and output and outcome data on the quality of service delivery—for example, service delivery measures. The following list provides a few examples of corruption along the production function:

1. **Inputs:** nepotism in hiring, procurement fraud, and off-budget leakages
2. **Processes:** unethical leadership by line managers and political pressures on public servants to act corruptly
3. **Culture and behavior in public administration:** whether public servants believe bribes are acceptable and corrupt behavior of public servants
4. **Outputs and outcomes in frontline agencies:** tax evasion in collusion with customs and tax officers
5. **Direct outputs and outcomes of public administration:** corruption in regulatory or policy decisions, bribery requests to citizens, and distorted allocation of licenses and permits.

FIGURE 8.1 Examples of Corruption along the Public Administration Production Function



Source: Original figure for this publication.

Note: The public administration production function conceptualizes the public sector as different dimensions (personnel, management practices, and attitudes and behaviors) that connect to each other to produce outputs and outcomes. HRMIS = human resources management information systems; IT = information technology; PFM = public financial management.

While our analysis focuses on each sector of the production function individually, integrating analytics across multiple dimensions of the production function provides even greater analytical insights. For example, procurement data may be linked to personnel data to identify public servants who might benefit from discretion over contracting decisions. Management practices, such as unethical leadership, may have downstream effects on the norms and behaviors of public servants. For this reason, integrating data allows practitioners and researchers to assess risk factors associated with corruption holistically, connecting the different parts of the production function. We include in each section a brief discussion of how to integrate different data sources and methodological approaches.

Each section of this chapter focuses on a particular dimension of the production function and how to measure and identify where corruption occurs. As such, the chapter should be read as an overview of analytical approaches to measuring corruption rather than as a corruption topic chapter. To frame our discussion, in each section we provide a brief overview of relevant academic and policy literature. We then highlight how methodological tools in other chapters of the *Handbook* can be used to understand and, potentially, reduce corruption. For example, if practitioners are interested in corruption in personnel, tools outlined in chapter 9, such as compliance and control mechanisms for human resources (HR) data, may prove helpful. We provide suggestions to practitioners about how to implement the methods discussed.

Accumulated experience suggests that relying solely on data is not enough to identify and measure corruption. Public officials and private agents who stand to benefit from corruption have incentives not to disclose their corrupt behavior in both survey and administrative data. As emphasized in chapter 4, measurement efforts are subject to existing power dynamics: politicians who benefit from collusion in procurement can manipulate procurement indicators to their advantage—for instance, by misreporting the number of bidders. Beyond concerns about the integrity of measures, improving the measurement of corruption should be embedded in a wider strategy of public sector reform to reduce corruption. A reform strategy can, for example, reduce corruption through monitoring technologies (for example, audits or reporting mechanisms) and positive reinforcement (for example, ethics training).³

Finally, we highlight the importance of combining data analytics with recognition of the sensitive political-economic issues surrounding reforms to reduce corruption (Evans 1995). Resistance to reform may come from multiple stakeholders who have economic incentives to preserve the status quo, not just from public officials. Politicians, political appointees, high-ranking public servants, and large corporations may resist data collection and analytics on corruption. Politicians may collude with or pressure public officials for personal gain, derailing reforms that threaten them.⁴ Survey data and interviews can help articulate the nature of these political dynamics and their intensity across the administration. Awareness of the institutional context within a country can guide reforms by securing buy-in from stakeholders and negotiating compromises that ensure the sustainability and effectiveness of reform efforts.

This chapter is structured as follows. Section 2 covers the input side of the production function, demonstrating how personnel, budget, and procurement data can be used to measure corruption. Section 3 dives into processes, such as management practices and business processes, that can be measured through a combination of survey and administrative data. Section 4 presents analytical approaches to measuring the norms and behaviors of public servants, particularly through surveys. Section 5 discusses corruption in frontline agencies, with a particular focus on service delivery, sectoral cases, and revenue collection. Section 6 covers the outputs and outcomes of public administration. Finally, we conclude.

INPUTS

Inputs to public administration include personnel, goods, capital, and budgets. In this section, we provide an overview of the extant literature on personnel (HRMIS and payroll), budget, and procurement, highlighting how different types of corruption, such as patronage, fraud, and embezzlement, may impair inputs into the public administration production function. Drawing on the approaches of other chapters in the *Handbook*,

we also present indicators to measure these dimensions of corruption and discuss how to develop and implement them.

Personnel and HRMIS Data

Personnel decisions, whether they regard selection or personnel management, have important consequences for public administration (Besley et al. 2022; Finan, Olken, and Pande 2017). Corruption may negatively affect personnel systems, particularly through patronage and nepotism, with long-term consequences (Evans and Rauch 1999; Rauch and Evans 2000). Patronage, the politically motivated selection and promotion of personnel, operates through the preferential hiring of copartisans (Brollo, Forquesato, and Gozzi 2017) or repayment for electoral contributions (Colonnelli, Prem, and Teso 2020). Patronage may adversely affect public servants' perceptions of the quality of governance and their general perceptions of corruption (Anderson, Reid, and Ryterman 2003). There can also be negative consequences for societal outcomes: the quality of government outputs such as health care and education can be compromised when people appointed based on political affiliation lack the skills or experience to perform critical functions.

It is important to consider not only how corruption occurs in personnel decisions (that is, through nepotism or patronage) but also how personnel systems affect corruption throughout public administration. A more meritocratic public service, for example, increases the opportunity cost for public servants who lose their jobs due to corruption (Cadot 1987). Conversely, if public servants believe that advancement is not based on good behavior, they may have an incentive to supplement their incomes through behavior that does not align with public policy goals (for example, bribes). Relatedly, if organizations are subject to high turnover due to political changes, officials will have a greater incentive to ensure their brief appointments pay off. Regarding the intensity of these political influences over bureaucratic careers, in Brazil's federal civil service, a quarter of civil servants believe that promotions are influenced by political connections (nepotism), and only 23.1 percent believe they are meritocratic (World Bank 2021). Such systematic prevalence of patronage may influence whether public servants engage in corruption.

Improvements in government analytics can assist in detecting and reducing corruption in personnel. Chapter 9 of the *Handbook* highlights a set of metrics that could be used to detect corruption in personnel. For example, talent management indicators that focus on recruitment—the number of applications per position or recruitment method (for example, competitive exam or political appointment)—can enable governments to better identify and measure cases of patronage. A lack of competitive exams or a low number of applicants may suggest a greater prevalence of patronage appointments. Performance indicators—the rate of performance reviews completed or employee ratings as completed by colleagues or supervisors—strengthen the measurement of meritocratic channels for promotion over political ones.

A publicly available analytics dashboard on public service can increase the transparency and accountability of personnel practices, as highlighted by the case study of Luxembourg in case study 1 of chapter 9. Moreover, HRMIS can enable governments to detect the risk of nepotism and patronage in recruitment by assessing similarity in the last names of public servants inside the same government organizations or, where such data are available, by linking name records to family records or political party membership records to understand where family or political party members are disproportionately hired into government (Bragança, Ferraz, and Rios 2015; Riaño 2021).

The implementation of analytical tools relies on robust data infrastructure, capacity, and attention to political-economic constraints. If HR data are to assist in the detection and reduction of corruption in personnel, governments need to establish processes for quality and compliance controls in HRMIS data to reduce gaps in the coverage and frequency of data. However, government agencies in which patronage is more common may resist sharing or even generating data on recruitment practices precisely to reduce this scrutiny. Additionally, there is the key issue of the sustainability of data collection. Governments often

launch new data collection efforts, but these efforts are not replicated over time. Collaborations with national statistical agencies and offices could ensure the sustainability of these efforts.⁵

Some countries do not have the necessary resources to implement an HRMIS. Thus, surveys and other tools are often used while an HRMIS is being designed and implemented. At the same time, it is important to complement efforts to generate HRMIS data on personnel with surveys of public servants and focus group discussions about experiences of corruption in personnel management. Political leadership from key stakeholders, such as the office of civil service, and broader institutional coordination may be necessary to reduce resistance by particular agencies.

Payroll

Corruption in payroll occurs through irregular payments to public servants, either through undue payments to ghost workers, who do not perform their service duties (Das and Hammer 2014; La Cascia et al. 2020), or through the collection of payments that exceed established guidelines (World Bank 2019). Payroll irregularities waste valuable public resources. In the context of fiscal constraint, irregular payments in excess of existing regulations may compromise the sustainability of the wage bill and lower citizens' trust.⁶ The irregular concentration of wages among a few public servants may lead to payroll inequalities that pose challenges to sustainability and may arise from public servants' wielding power to accumulate these resources (World Bank 2019). Reducing corruption in payroll is therefore an important policy objective for governments.

The principled use of payroll data, as well as the establishment of control and compliance mechanisms, can assist in curbing corruption in payroll. Chapter 9 outlines how indicators of payroll data can enable the accurate and timely monitoring of payroll in public administration. For example, a first exercise involves calculating the size of the wage bill and whether it complies with the actual budget allocation. A simple head count of public servants appearing in the payroll data identifies duplicated entries. A breakdown of wage-bill expenditures by sector, administrative unit, or territory enables a granular analysis of historical trends in payroll expenditure and a comparison between similar units to detect evidence of irregularity. Death and pension records can be cross-referenced with payroll data as well as attendance records to identify ghost workers. We note, however, that this exercise requires updated and reliable information systems. Compliance and control mechanisms, such as budget audits, should be set in place.

The use of digital technologies, such as machine learning, can assist in the detection of payroll irregularities, as outlined in chapter 16 and case study 2 of chapter 9. In particular, historical payroll data that are classified by payroll analysts as irregular can be used to train machine-learning algorithms to automatically classify irregularities in payroll entries. Given the large volume of payroll data being generated at any given point in public administration, these automated processes can complement and enhance the work of payroll analysts, enabling them to detect irregularities that would otherwise remain undetected. However, it is important to note that, in order to develop these advanced digital solutions, a robust payroll data infrastructure has to be set in place. Payroll data are often fragmented and decentralized. As chapter 9 outlines, a reform process may be necessary to integrate and standardize HRMIS data, which is demonstrated through its operational framework for HRMIS reforms.

The implementation of payroll reforms requires coordination with relevant stakeholders, such as the ministry of finance or the head of the civil service, and cognizance of the political-economic context. In particular, leadership support for reforms is necessary, as is navigating resistance from actors who benefit from the status quo, such as ghost workers. In contexts in which payroll data are highly decentralized, institutional coordination and data-sharing agreements are necessary to ensure that payroll data coverage improves. Additionally, an advisory rather than a punitive approach to payroll reform is recommended, in particular when justice systems are weak and unable to enforce regulations. While some duplications or excessive benefits may be intentional, they are often the result of a lack of knowledge or inadequate training, as well as legacy systems that are not regularly updated. As a result, onboarding to new control and compliance mechanisms, rather than punishing infractions outright, may reduce resistance to and ensure the uptake of policy reforms.

Budget

Budget data measure how public resources are spent throughout the entirety of the public administration. They include multiple sectors within the government, such as procurement and payroll. Due to its cross-cutting nature, the budget is exposed to different types of corrupt behavior. Corruption may manifest itself in budget leakages: resources that are unaccounted for in the flow of public resources across administrative levels of government (Gurkan, Kaiser, and Voorbraak 2009). In extreme cases, corruption occurs through embezzlement, the diversion of public funds for personal gain (Hashim, Farooq, and Piatti-Fünfkirchen 2020). Corruption in public expenditure may also have a distortionary effect, misallocating resources to less productive sectors of the economy and ultimately inhibiting economic growth (d’Agostino, Dunne, and Pieroni 2016). Because of its potential negative effects, corruption in the budget has been the subject of extensive policy debate, particularly in public financial management (World Bank 2020). Methodologies to detect corruption in budgets include the Public Expenditure and Financial Accountability (PEFA) assessment and Public Expenditure Tracking Surveys (PETS), which are discussed in greater detail in chapter 11.⁷

Chapter 11 also provides guidance on how to build more robust data infrastructures for public expenditures. It outlines five guiding principles that should be respected in designing and maintaining public expenditure data: data provenance and integrity, comprehensiveness, utility, consistency, and stability. These principles ensure that the sources of expenditure data are documented and accounted for and that data are comparable and stable across public administration. One simple measure is to map out all transactions in a given fiscal year to understand what goes through the government’s financial management information system (FMIS) and identify where high-value transactions are conducted. The share of the government budget transacted through the FMIS indicates the integrity of expenditure data.

Procurement

Governments are responsible for large volumes of procurement transactions. A recent estimate places these transactions at 12 percent of the global GDP (Bosio et al. 2020). There is a growing body of academic and policy literature on how to measure and prevent corruption in procurement. A widely used definition of corruption in procurement is the violation of impartial access to public contracts—that is, the deliberate restriction of open competition to the benefit of a politically connected firm or firms (Fazekas and Kocsis 2020). Corruption in procurement can occur in many forms. A single firm may bid for a procurement contract, securing exclusive access to lucrative government contracts. Or firms may overinvoice a procured good, often in collusion with procurement officials or politicians.

TABLE 8.1 Examples of Public Procurement Indicators

Economy and efficiency	Transparency and integrity	Competition	Inclusiveness and sustainability
<i>Tender and bidding process</i>			
<ul style="list-style-type: none"> • Total processing time • Evaluation time • Contracting time 	<ul style="list-style-type: none"> • Time for bid preparation • Single-bidder tender 	<ul style="list-style-type: none"> • Open procedure • Number of bidders • Share of new bidders 	<ul style="list-style-type: none"> • Share of SME bidders • Share of WOE bidders
<i>Assessment and contracting</i>			
<ul style="list-style-type: none"> • Awarded unit price • Final unit price after renegotiation 	<ul style="list-style-type: none"> • Share of excluded bids 	<ul style="list-style-type: none"> • Number of bidders • New bidders 	<ul style="list-style-type: none"> • Share of SME bidders • Share of WOE bidders
<i>Contract implementation</i>			
<ul style="list-style-type: none"> • Final unit price after renegotiation • Time overrun 	<ul style="list-style-type: none"> • Variation orders • Renegotiations 		

Source: Original table for this publication based on chapter 12.

Note: SME = small and medium enterprise; WOE = women-owned enterprise.

Chapter 12 provides an overview of a set of indicators and data sources on public procurement and how they can be used for data-driven decision-making (table 8.1). It also provides guidance on how to build data infrastructure and capacity for procurement data analytics and emphasizes the added value of combining public procurement data with other data sources. The chapter concludes by describing how a whole-of-government approach can increase the benefits of procurement analytics, as well as the advantages of combining administrative with survey data on procurement officials.

PROCESSES

Processes in public administration define organizational objectives and work procedures. They include management practices, which structure how managers and staff engage with each other in the performance of their duties. Processes also include business practices, which map onto the different regulations and procedures that structure how public servants should perform their duties. In this section, we provide a snapshot of the extant literature on these processes, highlighting, as illustrative examples, how unethical leadership and a lack of compliance with existing business processes may impact public administration. We also present indicators on these dimensions of corruption and discuss how to develop and implement them. Regarding data integration, management practices can affect multiple areas of public administration, including culture and behavior as well as turnover in personnel. It is therefore important to connect agency-level indicators of management practices with other administrative data sources.

Management Practices

Corruption in management practices involves the violation of business rules that govern how public servants are managed. We follow the definition proposed in Meyer-Sahling, Schuster, and Mikkelsen (2018), focusing particularly on human resource management practices. Management practices include decisions about recruitment into the public service, compensation, and the promotion of public servants. In practice, corruption can affect these different management functions. Politicians can appoint loyalists to the public service to extract rents, while low wages may encourage public servants to request bribes. Finally, political control or pressure may be applied to promote public servants who “steal for the team.”

Surveys of public servants can help measure their experience of (corrupt) management practices (see part 3 of the *Handbook* and also Meyer-Sahling, Schuster, and Mikkelsen 2018). In identifying these practices, practitioners must choose whether to capture public servants’ perceptions at the organizational level or their individual experiences with corruption. Chapter 23 assesses the trade-offs involved in each approach, highlighting how answers may differ depending on the kind of referent used. In particular, sensitive questions about topics such as corruption in management may be better measured at the organizational rather than the individual level because organizational comparisons reduce social-desirability bias. Another key question is how to assess the degree to which management practices differ across groups within public administration, such as across genders or organizations. Therefore, the choice of referent—organizational or individual—should be considered when designing the survey.⁸

Business Processes (Organizational Procedures in Government)

The *Handbook* provides tools to measure corruption in business processes, understood as the procedures that regulate how public administration is to conduct its business (for example, what kind of archives it needs to keep). Chapter 13 presents indicators that allow practitioners to measure the quality of the processes completed by public servants. Given that public administration is often regulated by a set of rules dictating which forms have to be filled out, the level of completion of these forms can be used by external evaluators to assess the quality of business processes in government and, thereby, the risk of corruption.

Indicators such as the availability of minutes, memos, and other relevant documents, as well as the proportion of incoming and outgoing correspondence with dates, stamps, and documented sources, may provide insights into the quality of business process implementation by public officials.

CULTURE AND BEHAVIOR OF PUBLIC SERVANTS

The culture of a public service organization includes the attitudes and norms of public servants, which, in turn, shape their behavior. Attitudes and norms include the values that guide officials as they do their jobs, the level of engagement officials have with their work, and the norms that govern their behavior, among other things. This section describes methods from the *Handbook* that can be used to assess how the culture, norms, attitudes, and behaviors of public servants might reveal a propensity for corruption. We have mentioned before that management practices can shape norms and behaviors in public administration. Integrating data on cultural norms with the quality of service delivery can help identify how norms shape practice in services such as health care and education.

Norms and Behaviors

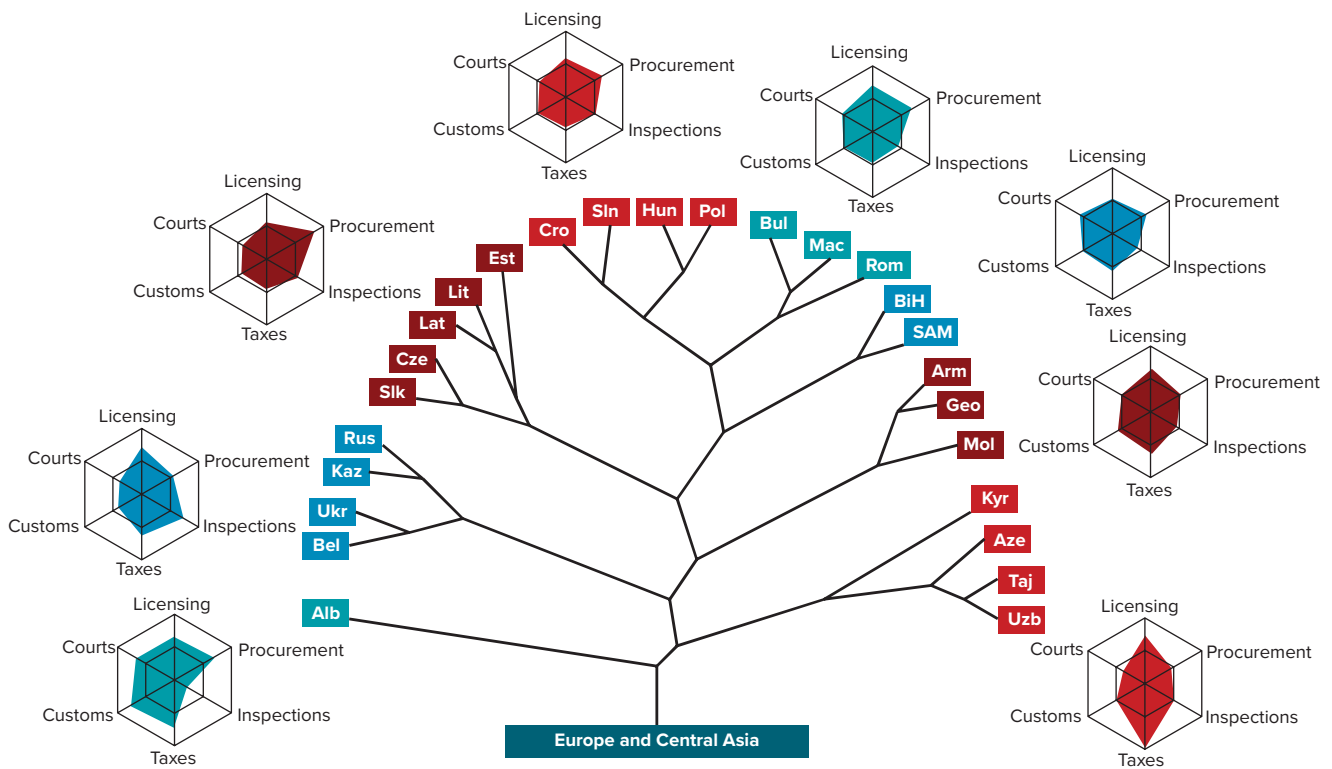
Corruption is most likely where the culture and norms within the public service enable it. The more prevalent corruption seems to public servants—the more it constitutes a norm—the more likely they may be to engage in corruption themselves (Köbis et al. 2015). Looking specifically at public servants, studies have found that certain motives, such as personal and social norms and opportunities not to comply with government rules, are significantly correlated with public servants' propensity to engage in corruption (Gorsira, Denkers, and Huisman 2018). These personal and social norms can also spur or hinder conflicts of interest, cases in which public servants' private interests unduly influence how they behave in public office, which have been a significant challenge in many countries. Initiatives led by various international organizations provide information on laws regulating conflicts of interest and their measurement.² However, business and ethical principles may vary across countries and within a single country. A report on Vietnam highlights different definitions of conflict of interest for public servants and how they can be culturally informed (World Bank 2016).

Identifying the attitudes and beliefs that motivate public servants to engage in corruption can help determine the root causes of corruption and inform strategies to curtail it at an international level as well. This task is crucial for practitioners examining corruption in the public service across countries, due to disparities in the understanding of corruption within different societies (Anderson and Gray 2006; World Bank 2016). Figure 8.2 shows clusters of countries based on the relative frequency of bribery in different sectors, enabling cross-country comparisons across different areas of corruption or bribes. These clusters map closely onto traditional groupings—for instance, northern and southern members of the Commonwealth of Independent States, Baltic states, and countries closest to accession to the European Union—suggesting that shared histories and similar paths of institutional development play a role in the types of corruption seen today.

The attitudes and motivations of individual public servants toward corrupt practices are primarily shaped by experiences and beliefs, making surveys a method well suited to analyzing them. However, self-reporting allows respondents to give inaccurate responses or not respond at all, distorting the resulting data. For example, if social or legal repercussions might arise from revealing a disposition toward corruption, public servants may try to mask their true attitudes or behavior. By applying methods in the *Handbook* to the design of surveys that aim to capture attitudes toward corruption, practitioners can mitigate distortions resulting from biases or nonresponse among public officials.

Chapter 22 presents findings about which questions can better solicit responses as well as a conceptual framework for understanding this phenomenon. One source of survey nonresponse is question sensitivity. A public servant asked about their views on violations of social norms or formally prohibited behavior may be hesitant to respond due to social-desirability bias or fear of legal sanctions. The chapter, however, suggests that

FIGURE 8.2 Country Clusters Based on Relative Frequency of Bribes in Specific Areas



Source: Adapted from Anderson and Gray 2006, figure 4.15.

Note: Alb = Albania; Arm = Armenia; Aze = Azerbaijan; Bel = Belarus; BiH = Bosnia and Herzegovina; Bul = Bulgaria; Cro = Croatia; Cze = Czechia; Est = Estonia; Geo = Georgia; Hun = Hungary; Kaz = Kazakhstan; Kyr = Kyrgyz Republic; Lat = Latvia; Lit = Lithuania; Mac = North Macedonia; Mol = Moldova; Pol = Poland; Rom = Romania; Rus = Russian Federation; SAM = Serbia and Montenegro; Slk = Slovak Republic; Sln = Slovenia; Taj = Tajikistan; Ukr = Ukraine; Uzb = Uzbekistan.

public servants *are* willing to answer sensitive questions on, say, their attitudes and behaviors toward corruption (though they may, of course, do so in a socially desirable way). Instead, the strongest predictor of nonresponse is complexity—specifically, a question’s “unfamiliarity” and “scope of information”—as when officials are asked general questions about the work environment rather than about their immediate experiences. To address this, survey questions should ask about public officials’ perceptions of corruption within their own organizations.

Merely eliciting a response, however, does not ensure that the data being collected through surveys are reflective of true norms surrounding corruption in public administration. For instance, to address whether face-to-face or online surveys better reduce response bias, chapter 19 examines the impact of survey mode on civil servant survey responses. Face-to-face surveys tend to offer several benefits, including significantly higher response rates and lower break-off rates. Online surveys, by contrast, limit the ability of an enumerator to probe public servants for responses, which risks distorting the true prevalence of attitudes and behaviors tolerant of corruption. Online formats tend to elicit more candid responses to potentially sensitive questions about topics such as ethics violations. Indeed, the chapter presents evidence that face-to-face surveys produce more “desirable” responses compared to online surveys—for instance, fewer public servants report that employees “observe unethical behavior among colleagues.” Survey designers must therefore consider the trade-offs of each survey mode, recognizing that the choice of survey mode can impact the accuracy of results. The pilot phase and focus group discussions can help validate survey results.

To draw comparisons regarding the norms that enable corruption, practitioners may compare survey results across different demographics, organizations, and countries. To do so, these different groups must understand survey measures in the same way. However, norms around corruption differ across countries and, at times, across organizations within a single country. Determining what public servants view as corruption or as an ethical violation is therefore necessary to understanding and contextualizing the results

of civil servant surveys.¹⁰ Chapter 24 provides an approach to this task by measuring the comparability of a latent statistical concept among different groups. While this chapter looks specifically at the concept of transformational leadership, its approach can be applied to latent concepts relating to attitudes toward corruption. By using this framework, practitioners can better ensure that when they measure corruption norms against certain benchmarks, those benchmarks enable valid comparisons.

Finally, due to the limitations of survey data and the potential rigidity of attitudes and norms regarding corruption, qualitative analyses can be valuable tools for interpreting the data obtained through surveys. Principles for using qualitative analyses to analyze norms and the behavior of public servants are presented in chapter 30. For example, through participant observation, observers can document how public servants interact in an office environment, which may be difficult to capture using a survey instrument. Meetings and other forms of interaction between public servants may reveal power dynamics—in particular, how gender, race, and seniority stratify hierarchies in public administration. Incorporating qualitative analyses such as these into investigations of corruption norms can give practitioners more robust insights into the roots of corruption and tools to remedy it.

DIRECT OUTPUTS OF PUBLIC ADMINISTRATION

Corruption may affect the outputs of governments in multiple ways. Politicians may exert pressure on public servants to relax procedures for electoral gain. Public servants may accelerate cases for firms in exchange for bribes. These forms of corruption can be measured through household and citizen surveys. This section draws on *Handbook* chapters to outline several illustrative indicators to measure corruption in case and task data, as well as citizens' perception of corruption. Since direct outputs are the products of public administration, these can be linked to multiple data sources, such as personnel, budget, and procurement. This enables practitioners to assess the efficiency of different personnel recruitment and management practices as well as norms and behaviors in producing important outcomes for citizens.

Case and Task Data

To implement policy, governments generate large amounts of administrative data on the deliberations and actions of public servants. These case data are rich records of how work is carried out within governments, and they enable practitioners and researchers to better understand public servants' performance within public administration. For example, exploring data on social security cases in Italy, Fenizia (2022) estimates the effect of management changes on office productivity, finding that improvements in manager talent increase productivity. Dasgupta and Kapur (2020) collect data on time-usage diaries in India to analyze bureaucratic task overload in block development offices. In other cases, public servants may face pressure from superiors to expedite bidding processes and cases or to disregard due process in the development of new regulations (World Bank 2021).

It is possible to detect fraudulent activity by public servants by analyzing case data. Chapter 15 provides a simple measurement, the error rate, to identify cases of potential fraud risk in public administration. Calculating the fraction of claims and cases processed incorrectly allows governments to measure the extent of administrative irregularities and provide remedial measures. For example, in the case of social security claims, there are two types of mistake: a government agency may incorrectly give beneficiaries a social transfer or erroneously deny a transfer to the correct party. Keeping track of appeals by denied beneficiaries would only capture the latter case. To provide a comprehensive measure of the error rate, and better identify fraudulent behavior, governments should regularly audit a random subset of claims by government offices.

Public servant surveys can also provide insight into the extent to which cases are affected by corruption. Chapter 13 provides indicators to assess the quality and completeness of claims in government through enumerator reviews of extant cases. To measure the comprehensiveness of reporting on a claim across a

series of tasks, the following survey question was asked: “Where applicable, are minutes, memos, and other necessary records present and complete?” Another indicator is the overall commitment to effective process: “In general, to what extent does the file organization adhere to government procedure?”

Household and Citizen Surveys

Household surveys are a valuable tool to measure corruption in public administration, generating evidence on citizens’ trust in government, the quality and accessibility of service delivery, and experiences with bribes (Anderson, Kaufmann, and Reanatini 2003; UNODC and UNDP 2018). A 1998 report on Latvia shows, using household surveys, that over 40 percent of households and firms agreed with the statement “A system whereby people could anonymously report instances of corruption would not be successful because corruption is a natural part of our lives and helps solve many problems” (Anderson 1998). Additionally, citizens reported that bribes often occurred in reaction to difficulties in processing cases with public servants, such as intentional delays to resolve firms’ requests or vague explanations of legal requirements. Numerous studies have found negative correlations between citizens’ perception of corruption in government and their level of trust and satisfaction with the government, two foundational components of effective political systems (Park and Blenkinsopp 2011; Seligson 2002).

Citizen surveys can assess the extent to which citizens trust their government, focusing on issues of corruption, such as bribery. To provide standardized metrics for measuring this relationship, chapter 28 develops a framework concerning drivers of trust in public institutions. This framework is based on four components of institutional trust, including a competence indicator that measures whether citizens believe public institutions “minimize uncertainty in the economic, social, and political environment” and a values indicator that measures whether public institutions “make decisions and use public resources ethically, promoting the public interest over private interests, while combatting corruption.” Disaggregating these components of trust allows practitioners to gauge both the level of corruption citizens expect from their government and the extent to which citizens believe corruption affects the reliability of public services. By applying this framework through the Organisation for Economic Co-operation and Development (OECD) Survey on Drivers of Trust in Public Institutions, as described in chapter 28, practitioners can examine the multidimensional effects of corruption on trust in public institutions and compare their results with other countries.

While corruption may impact trust in different ways across countries, a few practices can strengthen government transparency and accountability to mitigate the effects of corruption. Chapter 25 outlines several of these practices, which include disseminating public servant survey results across all levels of government as well as to the public. By disseminating survey results, which may include public servants’ perceptions of corruption within their agencies, governments can increase awareness of areas for improvement and incentivize stakeholders to act on results. One example comes from a recent corruption report in which a survey of federal civil servants in Brazil revealed that a majority had witnessed unethical practices while in the public service (World Bank 2021). By gathering and revealing this information, the comptroller general demonstrated his public commitment to addressing corruption.

FRONTLINE AGENCIES

Frontline agencies in government are responsible for the delivery of public services directly to citizens. These agencies include hospitals, schools, and police stations. Frontline agencies have been the subject of extensive research and policy work because their work has a direct impact on social welfare and their tasks, such as the delivery of health care or education, are often by nature amenable to measurement exercises. This section provides illustrative examples from the *Handbook* of how these agencies can be affected by corruption, such as bribery or budget leakages. Frontline agencies are amenable to different types of data integration:

personnel data help identify the extent to which service providers are qualified, management practice data provide information on how well services are being managed, and procurement data enable assessment of the extent to which important materials—for example, school textbooks or medical equipment—are sourced and distributed.

Service Delivery

Corruption may occur in different public services and sectors in the economy due to public servants' power to allocate benefits or impose additional costs on citizens, such as bribes (Rose-Ackerman and Palifka 2016; World Bank 2003). An example of this type of corrupt behavior is when a service provider—such as a government clerk issuing licenses or a health official providing services—extracts an informal payment from a citizen or business to grant or expedite access to the service (World Bank 2020). The World Bank Enterprise Surveys find evidence that, across the world, around 12.3 percent of firms are expected to give gifts to get an operating license.¹¹ Data from the World Bank Business Environment and Enterprise Performance Survey (BEEPS) suggest a higher prevalence of corruption in taxation and government contracts in contrast to utilities or environmental inspections (Anderson and Gray 2006). Collecting survey data on different respondents—for example, public servants, business owners, and citizens—paints a more holistic picture of corruption in service delivery. For example, the Governance and Anti-corruption (GAC) diagnostic surveys, developed by the World Bank, identify where a public servant asked for a bribe, or if a citizen first offered it.

In the *Handbook*, we provide different methodologies to measure corruption in both service delivery and sectoral data. For example, chapter 28 provides an indicator for measuring the integrity of public servants while they perform their duties, as observed by citizens. Specifically, the survey question is “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?” Asking citizens this question directly may reduce concerns over social-desirability bias that arise when surveying public servants, but both responses provide insight into how corruption in service delivery occurs.

Another indicator that may assist in measuring the prevalence of corruption is the time it takes to process social security cases, as outlined in chapter 15. While a delay in processing cases may not directly imply corruption, combining this information with citizens' perception of the prevalence of corruption may help identify particular sectors—such as social protection or business licenses—where delays in case processing are used as leverage to extract bribes from citizens.

Chapter 29 provides measures of service delivery (MSD) that can be used to identify cases of corruption in public service delivery. For health care in particular, patients may be asked questions regarding their confidence in the health care system, such as their level of satisfaction and recommendation, as well as regarding care uptake and retention. Low indicators for satisfaction may signal issues regarding public service delivery and can be followed by more direct questions about corruption. Additionally, indicators on the availability of medical equipment and supplies may help evaluate whether these resources are being allocated to relevant facilities or leaked along the way.

Finally, procurement indicators can be also used to evaluate contracts between a government's frontline agencies and contractors, as outlined in chapter 12. The prevalence of single-bidder contracts and short time periods for bid preparation point to problems in the bidding process, while a low number of renegotiations can shed light on collusion between government officials and contractors.

Regulation of Economic Sectors

Corruption is often linked to government officials' power to regulate what firms and individuals can do in particular sectors of the economy. Corruption in regulation affects environmental regulation, building inspections, labor and safety standards, and land-use management, among other areas. The rationale for government regulation emerged as a response to the rise of natural monopolies, such as in the telecommunications and energy industries, where the government should act in the public interest to reduce negative

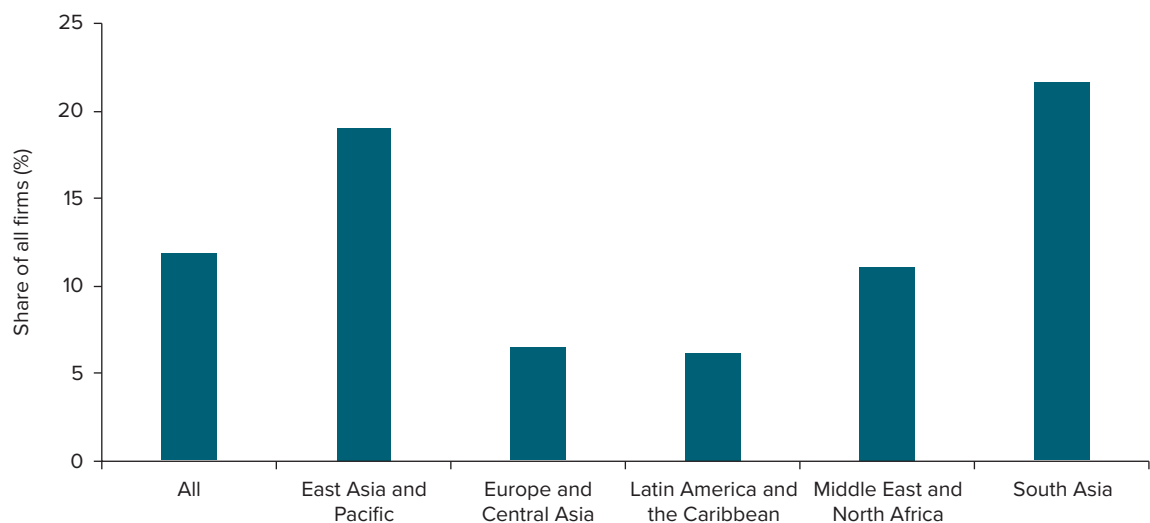
market externalities (Berg and Tschirhart 1988). However, regulation is subject to the pressure of interest groups and may be exposed to regulatory capture due to informational asymmetries and direct bribes (Laffont and Tirole 1991) or the promise of postgovernment employment in regulated industries (Dal Bó 2006). Because improving the informational environment can reduce the potential for regulatory capture, sectoral reforms have often focused on increasing transparency and access to data. A report on land management in Vietnam highlights how data can be used to track transparency and, in turn, how that transparency can reduce corruption in land-use regulation (World Bank 2014). The justice system plays an important role in enforcing regulation but can itself be exposed to corruption through bribery and other forms of capture.

The *Handbook* provides a few tools to understand corruption in sectoral cases, from both the supply and the demand side. On the demand side, chapter 28 highlights how measures of integrity and reliability can be applied to citizens’ experiences with regulatory agencies. For example, if a citizen is to apply for a business license, an indicator of the fairness of the process is the question “If you or a member of your family would apply for a business license, how likely or unlikely do you think it is that your application would be treated fairly?” Evaluating how business owners have been treated depending on age, education, and gender may yield additional insights about fairness. Questions regarding the integrity of public servants when adjudicating business license applications may yield insights into the extent to which regulatory agencies have been captured. Chapter 13 provides indicators on the supply side of government regulation. The degree to which internal government procedures are followed when adjudicating licensing and regulatory cases, as well as the degree to which cases are complete, provides evidence on whether governments are being impartial and thorough when adjudicating the enforcement of regulations on businesses and citizens.¹²

Revenue Collection

Corruption in revenue collection affects tax collection from citizens and revenue from customs, licensing fees, fines, and the sale of goods and services. Enterprise Surveys routinely point to tax as a source of corruption, often in the form of bribes to avoid paying the full tax amount, as illustrated in figure 8.3. This can weaken the tax base of the country and lead to government underfunding. Furthermore, corruption in tax administration can cause inefficiency by diverting resources from productive activity to bribery and

FIGURE 8.3 Percentage of Firms Expected to Give Gifts in Meetings with Tax Officials, by Region



Source: World Bank Enterprise Surveys, <http://www.enterprisesurveys.org>.

undermining equity, as honest taxpayers bear the brunt of these costs (Purohit 2007). In customs, corruption tends to take two forms: customs officials may pocket a portion of import revenue or may extract bribes from importers in exchange for some benefit (Yang 2006). Both forms of corruption can harm citizens by siphoning off revenue from the government, of which customs duties often constitute a significant share. Beyond financial costs, bribery in customs risks harming citizens by allowing illegal and potentially dangerous goods into the country. Through targeted incentives and heightened monitoring, corruption in customs can be curtailed (World Bank 2020).

Methods for mitigating corruption in revenue collection and administration are examined in chapter 15 using a case study from Punjab, Pakistan. The tools used to ensure the quality of tax administration include standardizing the process of reporting collected taxes and cross-checking the tax department's administrative records against the national bank's tax receipts for discrepancies. The chapter also assesses performance-pay mechanisms to improve revenue collection. Although these programs may increase revenue, they may also lead to greater bribes as performance rewards increase tax collectors' bargaining power over citizens. Governments should therefore consider the costs of potential incentive schemes for tax administrators.

Chapter 14 provides tools for identifying bottlenecks in customs processes, which lessen incentives to bribe officials to expedite the processes. These include the Automated System for Customs Data (ASYCUDA), time-release studies (TRS), trader perception surveys, and GPS trackers. Furthermore, the chapter describes revenue collection indicators that can be used to detect fraud. Where fraud is suspected, authorities can look at the value of identical goods or similar goods to determine the revenue that could have been collected had goods been correctly declared. Monitoring customs can be reinforced with mirror analyses that compare the quantity of goods declared by the exporting country to the quantity of similar goods declared by the importing country.

CONCLUSION

This chapter has provided a framework for both conceptualizing and measuring corruption in public administration. It recognizes corruption as a challenge with many faces and with particular characteristics and ways of operating in each sector of public administration. By assessing corruption along the public sector production function, we have highlighted the range of microdata approaches available in the *Handbook* to understand corruption in public administration. While our thematic focus has been on corruption, other topics in public administration are amenable to this approach as well. Of course, our review—drawing in particular on the chapters in this book—has only lightly engaged the vast literature on this topic. For example, hundreds of surveys on corruption in public administration have been conducted, with techniques ranging from list experiments to direct questions to randomized response techniques, to name a few.

Our goal in this chapter has been to highlight the different microdata sources and methodologies available to assess corruption and to show how measurement along the public administration production function can help analysts assess corruption holistically. This approach showcases the benefits of employing multiple sources of information in holistically assessing substantive topics in public administration, such as corruption. In so doing, we hope to have highlighted the benefits and challenges associated with the holistic use of government analytics in reducing corruption.

Beyond the comprehensiveness of analytics on corruption, a multipronged approach enables the integration of multiple data sources to reveal corruption in innovative ways. For example, integrating HRMIS with procurement data can reveal which procurement officials are more likely to manage corrupt bidding practices that benefit colluding companies. Additionally, information on wages, management practices, and other data sources can help diagnose what risk factors are associated with procurement officials' corrupt behavior. This integration of data provides exciting possibilities for understanding how corruption operates, generating novel insights that would not be possible if analytics were restricted to a single sector.

Finally, we have emphasized that measurement can provide direction for reducing corruption, but it does not reduce corruption itself. Challenging administrative reforms need to embed analytics of corruption into a broader public sector reform strategy, proactively using indicators to identify and enact policies against corruption. This requires not only technical knowledge but political negotiation based on compromise and consensus building. We hope that evidence-based reforms, informed by analytical insights, can guide practitioners in their efforts to understand and reduce corruption.

NOTES

We are grateful to Alexandra Jasmine Chau for excellent research assistance in completing this chapter. We are grateful for helpful comments from Daniel Rogger.

1. For a report on the multiple faces of corruption, see Campos and Pradhan (2007). For a comprehensive discussion of definitions of corruption, see Fisman and Golden (2017). Anderson et al. (2019) include a review of other definitions.
2. For a sample of innovative research and development on the use of data analytics to understand corruption, see video of the sessions from the Symposium on Data Analytics and Anticorruption (World Bank and Korea Development Institute School of Public Policy and Management 2021).
3. For a discussion, see chapter 4.
4. Focus groups and interviews enable involved actors to share their experiences with corruption and reveal these sources of pressure (Benjamin et al. 2014).
5. The experience of the National Institute of Statistics and Geography (INEGI), the national statistical agency of Mexico, sets an example for other countries about how to integrate data collection on corruption with regular data efforts. INEGI centralizes information on audits, experiences with corruption by citizens, and sanctions against civil servants. See the topic “Transparency and Anti-corruption” on the INEGI website: <https://en.www.inegi.org.mx/temas/transparencia/>.
6. This problem is highlighted in the Brazilian newspaper article “Problema não é número de servidores, mas salários altos, diz Temer a Bolsonaro,” *Folha de S.Paulo*, November 16, 2018, <https://www1.folha.uol.com.br/mercado/2018/11/problema-nao-e-numero-de-servidores-mas-salarios-altos-diz-temer-a-bolsonaro.shtml>.
7. More information about the PEFA assessment is available at the PEFA website, <https://pefa.org/>. For more information about PETS, see Koziol and Tolmie (2010).
8. A sampling tool for survey designers to assess the scale of sample required to investigate different topics in the public service is described in more detail in chapter 20 and can be found here: https://encuesta-col.shinyapps.io/sampling_tool/.
9. Examples of the measurement of conflict of interest include the Public Accountability Mechanisms (PAM) Initiative, EuroPAM, and the work done under the Stolen Asset Recovery (StAR) Initiative. Data on PAM are available at <https://datacatalog.worldbank.org/search/dataset/0040224>. Information about EuroPAM can be found on its website, <https://europam.eu/>. For more information about StAR, see World Bank (2012).
10. Anderson (2002) and World Bank (2002) also include direct survey questions—in the Kyrgyz Republic and Kazakhstan, respectively—about whether public servants define certain behaviors as “corruption.”
11. Data on corruption from the World Bank Enterprise Surveys are accessible at <https://www.enterprisesurveys.org/en/data/exploretopics/corruption>.
12. There are also de jure measures that look at gaps in the existing legal or regulatory structure that allow government officials to exercise discretion in favor of or against regulated entities. Mahmood and Slimane (2018) look at the existence of these structural weaknesses that allow firms to exercise privileges.

REFERENCES

- Anderson, James. 1998. “Corruption in Latvia: Survey Evidence.” Working Paper 33257, World Bank, Washington, DC.
- Anderson, James. 2002. *Governance and Service Delivery in the Kyrgyz Republic: Results of Diagnostic Surveys*. Washington, DC: World Bank. <http://documents.worldbank.org/curated/en/814871468046449300/Governance-and-service-delivery-in-the-Kyrgyz-Republic-results-of-diagnostic-surveys>.
- Anderson, James, David S. Bernstein, Jean Pierre Brun, Alexandra M. Habershon, Francesca Recanatini, Emile J. M. Van Der Does De Willebois, and Stephen S. Zimmermann. 2019. *Anticorruption Initiatives: Reaffirming Commitment to a Development Priority*. Washington, DC: World Bank. <http://documents.worldbank.org/curated/en/365421591933442799/Anticorruption-Initiatives-Reaffirming-Commitment-to-a-Development-Priority>.

- Anderson, James, and Cheryl W. Gray. 2006. *Anticorruption in Transition 3: Who Is Succeeding ... and Why?* Washington, DC: World Bank.
- Anderson, James, Daniel Kaufmann, and Francesca Recanatini. 2003. "Service Delivery, Poverty and Corruption—Common Threads from Diagnostic Surveys." Background Paper for the 2004 World Development Report Making Services Work for Poor People, World Bank, Washington, DC.
- Anderson, James, Gary Reid, and Randi Ryterman. 2003. "Understanding Public Sector Performance in Transition Countries: An Empirical Contribution." Working Paper 30357, World Bank, Washington, DC.
- Benjamin, Nancy, Kathleen Beegle, Francesca Recanatini, and Massimiliano Santini. 2014. "Informal Economy and the World Bank." Policy Research Working Paper 6888, World Bank, Washington, DC.
- Berg, Sanford V., and John Tschirhart. 1988. *Natural Monopoly Regulation: Principles and Practice*. New York: Cambridge University Press.
- Besley, Timothy J., Robin Burgess, Adnan Khan, and Guo Xu. 2022. "Bureaucracy and Development." *Annual Review of Economics* 14: 397–424. <https://doi.org/10.1146/annurev-economics-080521-011950>.
- Bosio, Erica, Simeon Djankov, Edward L. Glaeser, and Andrei Shleifer. 2020. "Public Procurement in Law and Practice." NBER Working Paper 27188, National Bureau of Economic Research, Cambridge, MA.
- Bragança, Arthur, Claudio Ferraz, and Juan Rios. 2015. *Political Dynasties and the Quality of Government*. Unpublished manuscript. <https://web.stanford.edu/~juanfrr/bragancaferrazrios2015.pdf>.
- Brollo, Fernanda, Pedro Forquesato, and Juan Carlos Gozzi. 2017. "To the Victor Belongs the Spoils? Party Membership and Public Sector Employment in Brazil." The Warwick Economics Research Paper Series (TWERPS) 1144, Department of Economics, University of Warwick, Coventry. <https://doi.org/10.2139/ssrn.3028937>.
- Cadot, Oliver. 1987. "Corruption as a Gamble." *Journal of Public Economics* 33 (2): 223–44. [https://doi.org/10.1016/0047-2727\(87\)90075-2](https://doi.org/10.1016/0047-2727(87)90075-2).
- Campos, J. Edgardo, and Sanjay Pradhan. 2007. *The Many Faces of Corruption: Tracking Vulnerabilities at the Sector Level*. Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/6848>.
- Colonnelli, Emanuele, Mounu Prem, and Edoardo Teso. 2020. "Patronage and Selection in Public Sector Organizations." *American Economic Review* 110 (10): 3071–99. <https://doi.org/10.1257/aer.20181491>.
- d'Agostino, Giorgio, J. Paul Dunne, and Luca Pieroni. 2016. "Government Spending, Corruption and Economic Growth." *World Development* 84: 190–205. <https://doi.org/10.1016/j.worlddev.2016.03.011>.
- Dal Bó, Ernesto. 2006. "Regulatory Capture: A Review." *Oxford Review of Economic Policy* 22 (2): 203–25. <https://doi.org/10.1093/oxrep/grj013>.
- Das, Jishnu, and Jeffrey Hammer. 2014. "Quality of Primary Care in Low-Income Countries: Facts and Economics." *Annual Review of Economics* 6 (1): 525–53. <https://doi.org/10.1146/annurev-economics-080213-041350>.
- Dasgupta, Aditya, and Devesh Kapur. 2020. "The Political Economy of Bureaucratic Overload: Evidence from Rural Development Officials in India." *American Political Science Review* 114 (4): 1316–34. <https://doi.org/10.1017/S0003055420000477>.
- Evans, Peter. 1995. *Embedded Autonomy: States and Industrial Transformation*. Princeton, NJ: Princeton University Press.
- Evans, Peter, and James Rauch. 1999. "Bureaucracy and Growth: A Cross-National Analysis of the Effects of 'Weberian' State Structures on Economic Growth." *American Sociological Review* 64: 748–65. <https://doi.org/10.2307/2657374>.
- Fazekas, Mihály, and Gábor Kocsis. 2020. "Uncovering High-Level Corruption: Cross-National Corruption Proxies Using Public Procurement Data." *British Journal of Political Science* 50 (1): 155–64. <https://doi.org/10.1017/S0007123417000461>.
- Fenzia, Alessandra. 2022. "Managers and Productivity in the Public Sector." *Econometrica* 90 (3): 1063–84. <https://doi.org/10.3982/ECTA19244>.
- Ferraz, Claudio, Frederico Finan, and Diana B. Moreira. 2012. "Corrupting Learning: Evidence from Missing Federal Education Funds in Brazil." *Journal of Public Economics* 96 (9–10): 712–26. <https://doi.org/10.1016/j.jpubeco.2012.05.012>.
- Finan, Frederico, Benjamin A. Olken, and Rohini Pande. 2017. "The Personnel Economics of the Developing State." In *Handbook of Economic Field Experiments*, edited by Abhijit Vinayak Banerjee and Esther Duflo, vol. 2, 467–514. Amsterdam: North-Holland. <https://doi.org/10.1016/bs.hefe.2016.08.001>.
- Fisman, Ray, and Miriam A. Golden. 2017. *Corruption: What Everyone Needs to Know*. Oxford: Oxford University Press.
- Gorsira, Madelijne, Adriaan Denkers, and Wim Huisman. 2018. "Both Sides of the Coin: Motives for Corruption among Public Officials and Business Employees." *Journal of Business Ethics* 151 (1): 179–94. <https://doi.org/10.1007/s10551-016-3219-2>.
- Gurkan, Asli, Kai Kaiser, and Doris Voorbraak. 2009. "Implementing Public Expenditure Tracking Surveys for Results: Lessons from a Decade of Global Experience." PREM Notes 145, World Bank, Washington, DC.
- Hashim, Ali, Khuram Farooq, and Moritz Piatti-Fünfkirchen. 2020. "Ensuring Better PFM Outcomes with FMIS Investments: An Operational Guidance Note for FMIS Project Teams." Governance Discussion Paper/Guidance Note Series 6, World Bank, Washington, DC.
- Jain, Arvind K. 2001. "Corruption: A Review." *Journal of Economic Surveys* 15 (1): 71–121. <https://doi.org/10.1111/1467-6419.00133>.
- Kaufmann, Daniel, Sanjay Pradhan, and Randi Ryterman. 1998. "New Frontiers in Diagnosing and Combating Corruption." PREM Notes 7, World Bank, Washington, DC.

- Köbis, Nils C., Jan-Willem van Prooijen, Francesca Righetti, and Paul A. M. Van Lange. 2015. “Who Doesn’t?—The Impact of Descriptive Norms on Corruption.” *PLoS One* 10 (6): e0131830. <https://doi.org/10.1371/journal.pone.0131830>.
- Kozioł, Margaret, and Courtney Tolmie. 2010. *Using Public Expenditure Tracking Surveys to Monitor Projects and Small-Scale Programs: A Guidebook*. Washington, DC: World Bank. <http://documents.worldbank.org/curated/en/388041468177874064/Using-public-expenditure-tracking-surveys-to-monitor-projects-and-small-scale-programs-a-guidebook>.
- La Cascia, Hunt, Izza Akram Malik, Eduardo Vicente Goncalves, Yasodara Maria Damo Cordova, and Michael Kramer. 2020. *Finding Fraud: GovTech and Fraud Detection in Public Administration*. Washington, DC: World Bank.
- Laffont, Jean-Jacques, and Jean Tirole. 1991. “The Politics of Government Decision-Making: A Theory of Regulatory Capture.” *The Quarterly Journal of Economics* 106 (4): 1089–127. <https://doi.org/10.2307/2937958>.
- Mahmood, Syed Akhtar, and Meriem Ait Ali Slimane. 2018. *Privilege-Resistant Policies in the Middle East and North Africa: Measurement and Operational Implications*. MENA Development Report. Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/29353>.
- Meyer-Sahling, Jan-Hinrik, Christian Schuster, and Kim Sass Mikkelsen. 2018. *Civil Service Management in Developing Countries: What Works? Evidence from a Survey with 23,000 Civil Servants in Africa, Asia, Eastern Europe and Latin America*. Report for the UK Department for International Development (DFID).
- Park, Heungsik, and John Blenkinsopp. 2011. “The Roles of Transparency and Trust in the Relationship between Corruption and Citizen Satisfaction.” *International Review of Administrative Sciences* 77 (2): 254–74. <https://doi.org/10.1177/0020852311399230>.
- Purohit, Mahesh C. 2007. “Corruption in Tax Administration.” In *Performance Accountability and Combating Corruption*, edited by Anwar Shah, 285–302. Washington, DC: World Bank.
- Rauch, James E., and Peter B. Evans. 2000. “Bureaucratic Structure and Bureaucratic Performance in Less Developed Countries.” *Journal of Public Economics* 75: 49–71. [https://doi.org/10.1016/S0047-2727\(99\)00044-4](https://doi.org/10.1016/S0047-2727(99)00044-4).
- Recanatini, Francesca. 2011. “Assessing Corruption at the Country Level.” In *Handbook of Global Research and Practice in Corruption*, edited by Adam Graycar and Russell Smith, 34–64. Cheltenham: Edward Elgar.
- Riaño, Juan Felipe. 2021. “Bureaucratic Nepotism.” STEG Working Paper, Research Theme 5: Political Economy and Public Investment, Structural Transformation and Economic Growth, UK Foreign, Commonwealth & Development Office. <https://steg.cepr.org/publications/bureaucratic-nepotism>.
- Rose-Ackerman, Susan. 1978. *Corruption: A Study in Political Economy*. New York: Academic Press.
- Rose-Ackerman, Susan, and Bonnie J. Palifka. 2016. *Corruption and Government: Causes, Consequences, and Reform*. 2nd ed. Cambridge: Cambridge University Press.
- Seligson, Mitchell A. 2002. “The Impact of Corruption on Regime Legitimacy: A Comparative Study of Four Latin American Countries.” *Journal of Politics* 64 (2): 408–33. <https://doi.org/10.1111/1468-2508.00132>.
- UNODC and UNDP (United Nations Office on Drugs and Crime and United Nations Development Programme). 2018. *Manual on Corruption Surveys: Methodological Guidelines on the Measurement of Bribery and Other Forms of Corruption through Sample Surveys*. Vienna: UNODC. https://www.unodc.org/documents/data-and-analysis/Crime-statistics/CorruptionManual_2018_web.pdf.
- World Bank. 2002. *Kazakhstan Governance and Service Delivery: A Diagnostic Report*. Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/15098>.
- World Bank. 2003. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank.
- World Bank. 2012. *Public Office, Private Interests: Accountability through Income and Asset Disclosure*. Stolen Asset Recovery (StAR) Series. Washington, DC: World Bank.
- World Bank. 2014. *Land Transparency Study: Synthesis Report*. Washington, DC: World Bank. <http://documents.worldbank.org/curated/en/934431468121474463/Land-transparency-study-synthesis-report>.
- World Bank. 2016. *Managing Conflict of Interest in the Public Sector: Law and Practice in Vietnam*. Washington, DC: World Bank. <http://documents.worldbank.org/curated/en/580601478253123042/Managing-conflict-of-interest-in-the-public-sector-law-and-practice-in-Vietnam>.
- World Bank. 2019. *Anticorruption Initiatives: Reaffirming Commitment to a Development Priority*. Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/34010>.
- World Bank. 2020. *Enhancing Government Effectiveness and Transparency: The Fight against Corruption*. Washington, DC: World Bank. <http://documents.worldbank.org/curated/en/235541600116631094/Enhancing-Government-Effectiveness-and-Transparency-The-Fight-Against-Corruption>.
- World Bank. 2021. *Ethics and Corruption in the Federal Public Service: Civil Servants’ Perspectives*. Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/36759>.
- World Bank and Korea Development Institute School of Public Policy and Management. 2021. “Symposium on Data Analytics and Anticorruption.” Filmed October 25–28, 2021, at the World Bank, Washington, DC, and online. Video, 3:50:10. <https://www.worldbank.org/en/events/2021/10/25/symposium-on-data-analytics-and-anticorruption#1>.
- Yang, Dean. 2006. “The Economics of Anti-Corruption: Lessons from a Widespread Customs Reform.” In *International Handbook on the Economics of Corruption*, edited by Susan Rose-Ackerman, 512–46. Cheltenham: Edward Elgar.