PART II

Understanding Patterns of Reform

3
What Does Reform Look Like? Mapping Reform Efforts over Time

This chapter, like the book as a whole, focuses mainly on describing big-picture patterns in reform design and implementation that repeatedly recur across countries and time periods. But before we look at these broader patterns, it might be helpful to delve into the story of a single reform: Zambia's Public Service Capacity Building Program (PSCAP), which began in 2000. PSCAP was neither remarkably successful nor unsuccessful, but it did exemplify several of the most common features of reforms around the continent.

To understand PSCAP, we first need to know a bit about the context in which it was designed. PSCAP was Zambia's second major effort at civil service reform since the country's return to multiparty democracy in 1991. Zambia's first effort—the Public Service Reform Programme (PSRP, 1993–1999)—had been shaped by two linked imperatives for the new government: first, the need to reduce the government wage bill in the context of fiscal retrenchment and structural adjustment; and second, the perception by President Frederick Chiluba's democratically elected administration that the existing civil service was bloated by unqualified patronage hires of former President Kenneth Kaunda under the preceding decades of single-party rule. These two imperatives together shaped the goal of the PSRP to create a "more efficient but smaller public sector."[1] This was to be achieved by two main strands of central government reforms: (1) laying off 25 percent of civil servants but imposing higher education requirements and improving pay and conditions for those who remained and (2) "link[ing] pay and performance in a way that would attract and retain skilled professionals in the civil service."[2]

The PSRP achieved neither of these goals, though it did take some steps in each direction. Many staff were laid off, but fewer than envisioned, and the actual cost savings were minimal. This was due to a combination of resistance from unions, the political pain of imposing mass redundancies, costly retrenchment payouts and lawsuits, and allegations of partisan bias in who was laid off. Even when older workers appointed by the Kaunda administration were fired, there was a widespread perception that (despite many being patronage appointees) these experienced staff actually had more practical know-how than the younger workers with more formal education who replaced them. As one civil servant working during that time explained, "We ended up with people that were qualified, but surprisingly not competent."[3] A new staff appraisal system was created to link performance to pay and promotion, but it was no more than a formality, with no actual rewards or sanctions attached. Other redundancies—both forced and voluntary—were targeted at "nonessential" staff, which resulted in hospitals that had surgeons but no mortuary attendants or staff to clear operating rooms, as well as boarding schools with no cooks. This caused service delivery to "hit a disaster level, especially when it came to the frontline services like health and education."[4]

PSCAP was born of these perceived failings. PSRP had been driven by the assumption that bringing in more qualified staff with better salary structures would improve civil servants' performance, but when this did not materialize, it was argued by some (including many civil servants) that perhaps they lacked the necessary equipment and resources to improve. PSCAP's mantra and core goal were, therefore, to improve individuals' and institutions' "capacity to deliver" (hence the emphasis on *capacity* in its name).[5] PSCAP had a total anticipated cost of US$45 million and was funded by a World Bank project loan of US$28 million, US$16 million from the UK Department for International Development (DFID) and other donors, and US$1 million from the government of Zambia.

The main vehicle for directly trying to increase capacity was a Performance Improvement Fund (PIF), to which service delivery organizations in priority areas could apply to fund discrete, quick-win projects to demonstrate that reforms could yield tangible results. PIF applications were formulated by the organization itself, and this bottom-up process led to several innovative ideas for projects such as mobile hospitals and mobile education labs, including a UN award for a grant addressing the scarcity of medical facilities for deliveries in the country's Copperbelt region. These PIFs were intended to be embedded within the broader organizational strategic plans that began to be developed under the PSRP, serving as a small-scale accelerant and demonstration of success.

But despite these initial successes, PIFs did not catalyze the broader impact they were intended to. Whereas the funds were meant to be targeted toward seven priority service delivery organizations, the World Bank reported that "in the early stages of the project, PIF funds were made widely available to all restructured ministries, regardless of whether they had a direct public service delivery orientation or not."[6] After a refocus on service delivery, an internal review of the PIF found that only fifteen of thirty-eight projects funded under the PIF had "an observable impact on service delivery," indicating "a disconnect between the service delivery and strategic planning basis for PIF funded projects and the projects actually funded."[7] Of the innovations introduced, reportedly only the mobile labs in education were sustained after the end of the project funding. Another study found that the PIF: "has not lived up to expectations. The logic of PIFs and quick wins was not sufficiently embraced by the MDAs [ministries, departments, and agencies]. PIFs were seen as supplementary financing to government allocations. As a consequence, most applications for PIF funding were inappropriate (for example, cars and computers), lacked both innovation and a focus on performance improvement, and were not linked to MDAs strategic plans."[8]

The perception that the PIFs became viewed mainly as a way to purchase equipment gave rise to a joke I heard from several people about PSCAP's acronym: that it stood for "please sir, can I have another Pajero," in reference to the 4×4 vehicle model popular with officials of government agencies (as well as with donor and NGO staff). At the same time, the severe spending restraints in place meant that there were real shortages of equipment within the government. As one former civil servant involved in designing and implementing the reforms explained, ministries would approach the implementing secretariat and say, "We are incapacitated because we don't have a vehicle," so there was actually some service delivery rationale for these purchases—even if there were also distortions of the PIF's intent.[9]

Alongside its organizational performance component, PSCAP also undertook a set of staffing and pay reforms following the PSRP. In reaction to the growing unpopularity of layoffs, PSCAP continued the organizational restructuring and associated retrenchments but reframed this as "rightsizing" rather than downsizing and did increase hiring in some areas, particularly in frontline social service delivery roles. Despite this rhetorical change of tack, staff numbers continued to decrease, with a net reduction of twenty-four thousand staff between 2000 and 2003. Although there were some hiring increases for teachers and nurses, ongoing fiscal challenges with overall payroll figures meant that staff strength was sometimes increased on paper (through notional approval for higher staff numbers), but financial clearance was not given to actually hire people, resulting in positions being left unfilled for many years.

At the same time, PSCAP also reiterated the PSRP's goal of establishing performance-linked incentives for individual civil servants. The government's *2002 Medium-Term Strategy for Enhancing Pay and Conditions of Service* ("the

Valentine report") was unequivocal about this: "The newly articulated pay policy should as much as possible, aim at explicitly linking pay to performance, signaling a major change in the incentive system and in performance expectations. Rewards and penalties are both vital for a well functioning incentive regime. . . . Meaningful performance incentives are a must." [10]

The government aimed to achieve this goal of instituting performance-linked incentives through two channels: the proper implementation of the Annual Performance Appraisal System (APAS) introduced during the latter stages of the PSRP reform for rank-and-file civil servants up to the level of director and the creation of performance contracts for permanent secretaries (bureaucratic heads of ministries).

With respect to APAS, PSCAP did not change the formal design of the system substantively but sought to actually attach incentives—such as differential pay increments, accelerated or delayed promotion, and meaningful sanctions—to the results of officers' APAS appraisals. This effort was not successful. External reviews in both 2005 and 2008 reported that there were no rewards or sanctions attached to the results of the APAS appraisal, with one report stating that "good performance is not rewarded while poor performance goes unpunished" and remarking that many employees do not even go through the appraisal process on an annual basis.[11] These faults are blamed not on the system itself, which is "adequately designed," but on the system's implementation and "low commitment of its users."[12] The second evaluation's overall assessment was that the APAS system "results in little individual or organizational performance improvement," and was devastatingly frank about the situation: "As time has passed the real purpose of the APAS report has become the justification of pay increments and promotions. This has led to the a [*sic*] view that completion of the form is a necessary evil to which one should devote as little time and thought as possible. The result in many instances is a report replete with inconsistencies, contradictions and very little assessment of performance that bears little relation to a real work plan and virtually none to the organisational and strategic plan."[13]

For more senior bureaucrats (permanent secretaries) who were not subject to the APAS system, PSCAP introduced a system of performance contracts that were intended to establish and measure organizational performance targets and link the renewal of permanent secretaries' contracts to these formal assessments. However, what ended up happening was that permanent secretaries were put onto fixed-term three-year contracts at the end of 2001 (rather than the permanent and pensionable civil servants they had been) but without meaningful setting of targets, assessments, or incentives. Permanent secretaries were willing to accept the temporary contracts because they promised a lucrative three years prior to retirement, and the arrangement was also "user-friendly" to politicians in that it gave them greater discretion and leverage over permanent secretaries.[14] However, permanent secretaries lacked not only annual targets but also basic job descriptions, so in practice, there was no formal linkage between performance and incentives. These problems were easily foreseeable, with a World Bank–funded consultancy reporting in 2002: "It is unclear at this point on what basis performance will be measured. What benchmarks will be used to objectively distinguish between levels of performance, particular[ly] since the MDAs have not completed the strategic planning process [that was also a component of PSCAP] and thus do not have clear performance targets."[15]

Taken together, these major components of the PSCAP reform—the Performance Improvement Fund, "rightsizing" redundancies, pay reform, the institutionalization of the APAS appraisal system, and the creation of performance contracts for permanent secretaries—fell short of the transformative impact on service delivery that they envisioned. The overall sentiment regarding the limited performance impacts of the PSRP and PSCAP reforms is well-captured by a 2008 review of Zambia's linked performance management reforms to date, which found that they resulted in "only marginal impact on the effectiveness and efficiency of the public service and result[ed] in little individual or organizational performance improvement."[16] PSCAP had initially been envisioned as a three-phase, thirteen-year program running from 2000 to 2013. As a result of these perceived shortcomings, however, PSCAP was terminated in 2005 at the end of phase one.

But PSCAP was only one reform among many in Zambia, and Zambia is only one of six countries whose reform histories are examined in this book. Let's step back now and look at the big picture of reform activity across all these countries based on the reform mapping described in chapter 1 and the appendix.

PATTERNS OF REFORM

This section asks and answers three descriptive questions about reforms:

1. How frequent have reform efforts been?
2. What has been the content of reforms (i.e., the structures, processes, or practices they have tried to introduce)?
3. How successful have reforms been?

How Frequent Have Reform Efforts Been?

Here's the simple answer to this question: There have been *a lot* of reform efforts, and they overlap with one another both in timing and substance. That said, defining and counting reforms is surprisingly difficult to do precisely. I'll talk through the evidence underlying this answer and discuss the challenges in quantifying it.

Figure 3.1 presents timelines of reform for each of the six countries. The top row lists the party and president in office during each year, the bottom rows list all the reforms that fit my definition that were underway in each country at each point in time, and the middle row lists what reform era (or eras) the many distinct reforms ongoing at any point in time can best be grouped into. Full country-by-country narratives of these reform histories are contained in the appendix.

In each country, there was almost always a reform underway in each year—usually several. We sometimes assume that civil service reforms are occasional upheavals with long periods of stagnation or that they are only initiated when political conditions are propitious. That might be true of the highest-profile reforms, but taking the approach of mapping out the universe of reforms—including minor, abortive, or little-known reforms—reveals that, in reality, new reform efforts were constantly being initiated in every country.

What's more, these reforms overlapped with one another in several ways, and the boundaries between them were unclear. For example, in Zambia, PSCAP was clearly the main reform ongoing between 2000 and 2005. Some of the pay and salary structure reforms were folded into PSCAP, while others—such as the Valentine report on pay policy and the creation of performance contracts for permanent secretaries—were not formally part of the PSCAP project but were designed and implemented with awareness of PSCAP and an aim to ensure coherence between them. Should they be counted as three different reforms or one? Or should the semi-independent subcomponents of PSCAP, such as the PIF and operationalization of the annual Performance Appraisal System (APAS), be separated from PSCAP as a whole and counted as reforms in their own right? In other words, did Zambia have one reform ongoing between 2000–2005, or three, or more? Should government-led reforms be counted as separate from the donor support projects that helped fund them—and does it matter if they have the same or different names or if the start and end dates align? Just looking at these timelines reveals not only the degree to which different reform efforts overlap with one another temporally (and thus are linked in planning and execution) but also the nonoverlaps of start and end dates (which hint at the less-than-full coherence of these efforts).

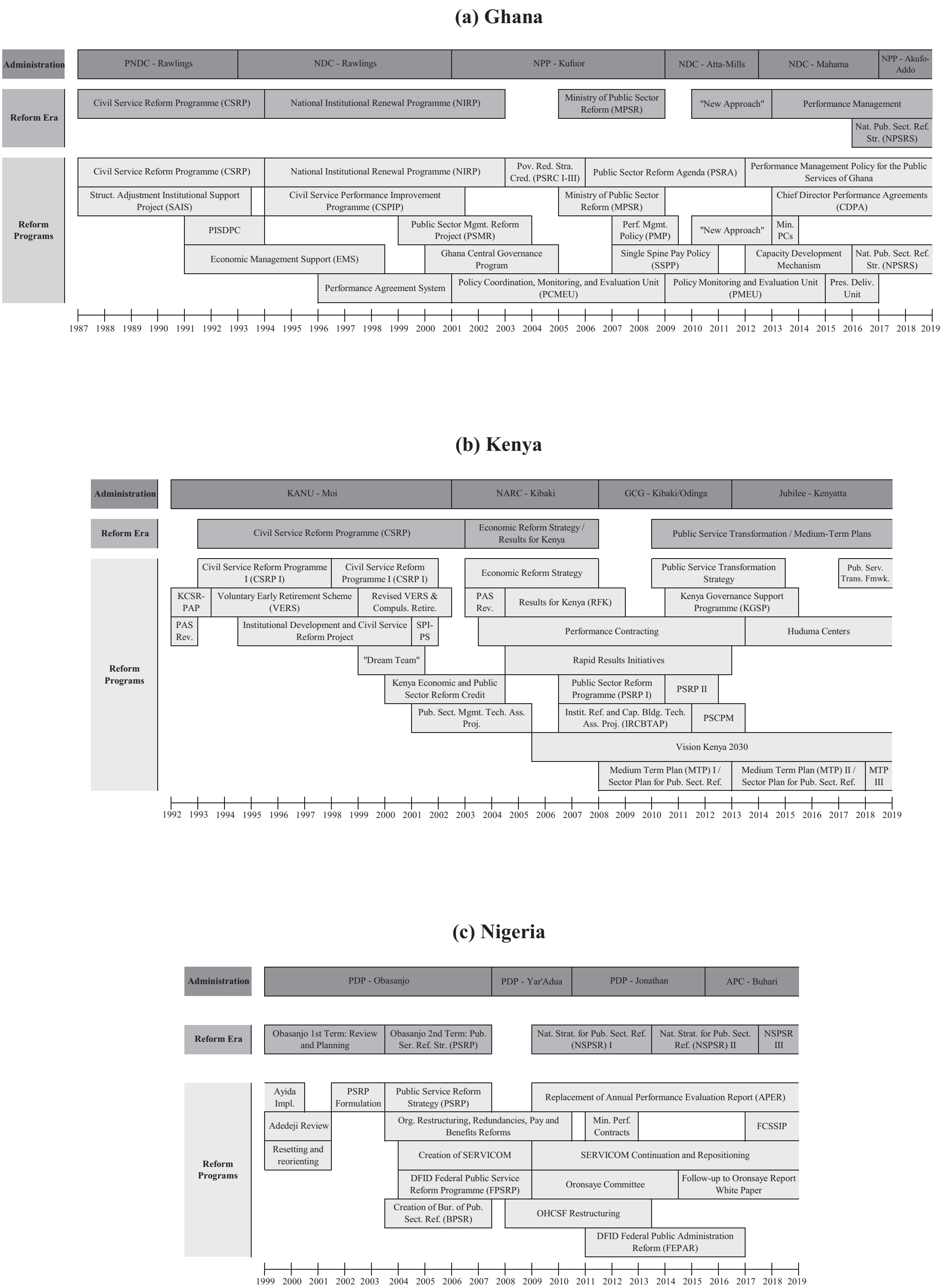

FIGURE 3.1 Timelines of Reform Efforts in Six Countries

Source: Author's synthesis. See appendix for more details and full names of each reform.

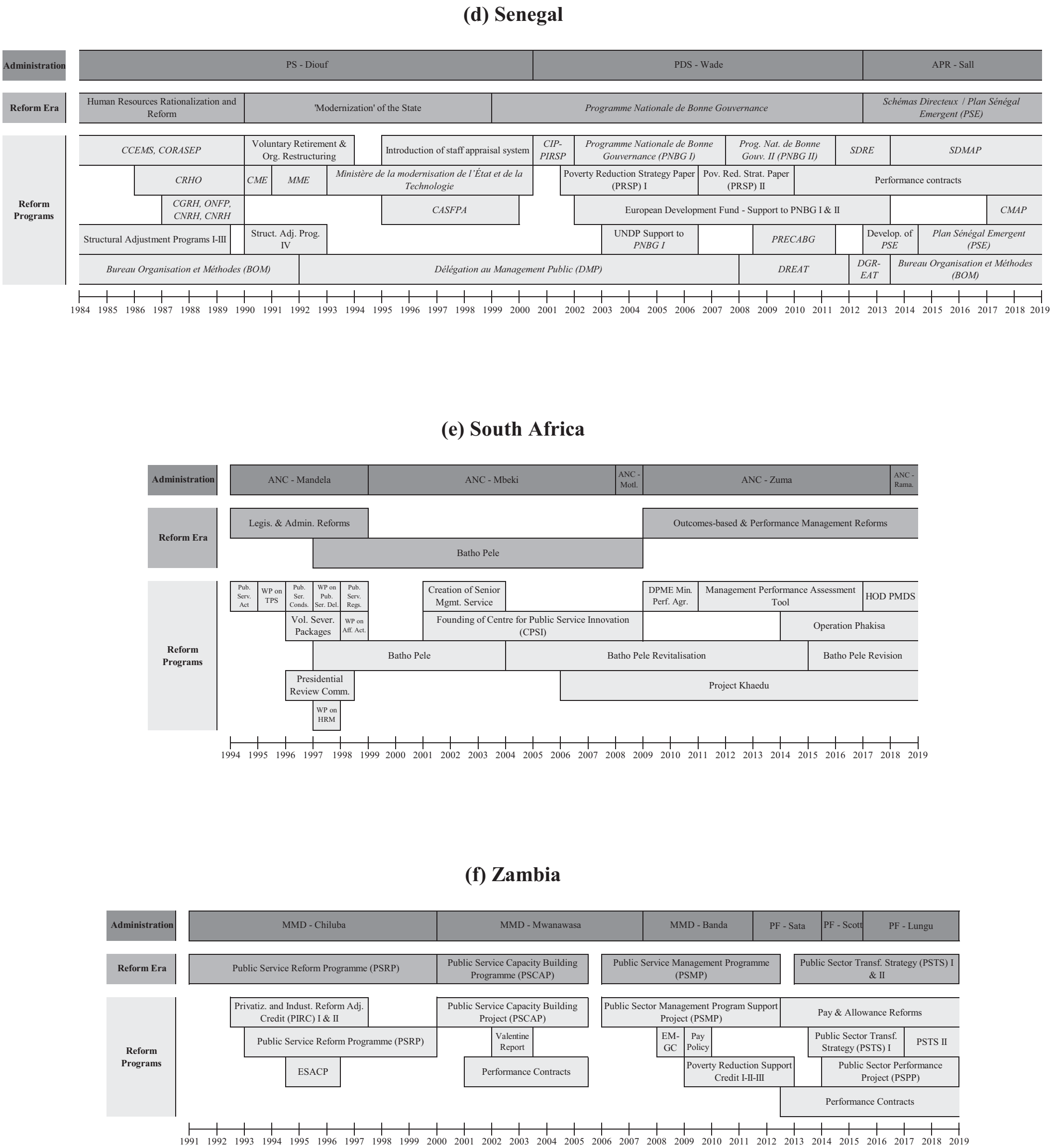

FIGURE 3.1 (*continued*)

The boundaries between reforms are blurry in another way: Each reform was influenced by the reforms that came before it. Take the example of Zambia's PSRP (1993–1999) and PSCAP (2000–2005). In some instances, this path dependence reflected reformers' efforts to build on prior reforms. In Zambia, for example, PSRP designed and introduced the APAS system; PSCAP then tried to make it function as envisioned. But in other instances, reformers did the opposite and deliberately tried *not* to connect their efforts to what had come before in an effort to avoid their perceived failings and being associated with them. Indeed, PSCAP's core approach and rhetoric—focusing on supporting bureaucrats' capacity to deliver—was deliberately chosen in reaction to the widespread perception that PSRP had negatively affected service delivery. Other elements of PSCAP combined both types of path dependence, such as the continuation of PSRP's efforts to rationalize staff numbers and allocation but with the rebranded goal of "rightsizing" rather than "downsizing." Thus, while PSRP and PSCAP are clearly distinct reform efforts in many ways, they are also inextricably linked. So no matter how one decides to aggregate the various reforms ongoing in a country at any given point in time, any analysis of these reforms has to take into account that none of them are actually independent from the others.

There is also the question of where to draw the line in terms of when a reform comes into existence. Is it when it gets an official plan and acronym, when a press conference announcing it is held, when the first internal brainstorming meeting happens, or when a particular idea first crosses someone's mind? And how should we treat reforms that got discussed but, for various reasons, were never formally adopted? For example, one civil servant in Ghana told of an idea to create a senior management service in the mid-2000s that was discussed and researched internally but never saw the light of day due to opposition from the civil service union.[17] Similarly, some types of reform efforts—those that generate news coverage, official documents, and academic studies—are more likely to be recognized as reforms both by researchers as well as public servants themselves and thus may be more likely to appear in the data. These measurement questions are further complicated by the possibility that the same factors that determine whether a reform gets launched might also be related to its success, so from a causal inference perspective, even the existence of a reform case to analyze is endogenous. So the sample of reforms that we have to count and study is almost certainly biased in many ways despite the numerous methodological steps I took to recognize and minimize the extent of this bias (detailed in the appendix).

In compiling reform histories, these dilemmas repeated themselves across each country and each time period, as there were multiple interpretations possible of the definition and measurement criteria I set out in chapter 1. In the end, I opted for a balance between consistency of approach across all six countries and representing reform histories in a way that would be qualitatively recognizable to individuals who worked in each civil service during these reforms. Representing both individual reforms (rectangles in the bottom rows of figure 3.1) and overall reform eras (rectangles in the middle rows) on the timeline also hints at the multiple possible ways of viewing these reforms. All in all, from an analytical perspective, these reform efforts are best regarded as semi-independent cases with blurry boundaries between them, which makes it problematic to attempt precise quantitative coding or analysis of the data.

With those caveats in mind, it's still useful to give some kind of numerical answer to the question of how frequent reforms are. The simplest and most true-to-reality way to do this is to count the number of individual reform efforts (bottom row rectangles) represented on the timeline. Doing so yields a figure of 131 reforms across the 173 country-years captured on the timeline. That equates to a new reform effort being launched in each country once every 1.3 years, on average. In some ways, this actually understates the amount of reforms undertaken since many of these efforts are actually bundles of multiple changes packaged under the same banner—as the next section discusses.

Figure 3.2 shows the distribution of the number of active reforms that were simultaneously ongoing during all country-years in the data—the number ranged from zero up to nine.[18] The modal number of reforms active in any given year across the whole sample was four.

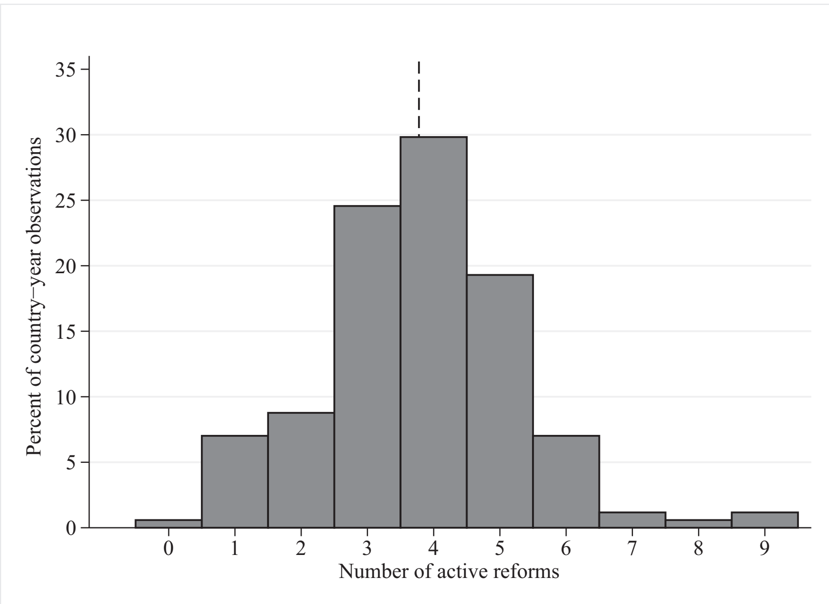

FIGURE 3.2 Number of Active Civil Service Reforms Per Country-Year

Source: Author's calculations.

Choosing different ways of aggregating and linking reforms could give lower (or even higher) answers. But by any measure, civil service reform efforts in these countries were very frequent, with significant (but not perfect) overlap with one another.

What Has Been the Content of These Reform Efforts?

These reform efforts encompassed a wide range of attempted changes to structures, processes, and practices in their respective civil services. But the pattern that emerges most clearly from an analysis of the content of these reforms is that governments frequently tried to implement similar types of reforms over and over again.

To see this, let's begin by looking at the content of the reforms implemented across the four major eras of reform in Zambia, as represented by table 3.1. For each cluster of reforms, I classified if it attempted to introduce each of ten different types of change, split across five different categories.

The first content category, individual-level performance management, captures reforms that tried to link some measure of individual performance to some type of extrinsic reward or sanction. Individual-level performance management reforms fall into two categories: those aimed at rank-and-file civil servants (like Zambia's APAS) and those aimed at senior leadership and managers (like Zambia's performance contracts under PSCAP).

The second category, organizational management or capacity, captures reforms that aimed to use organizational reviews, routines, or resources to improve collective performance at an organizational level. One subcategory of these was performance improvement funds established for specific, demand-driven reform activities—like Zambia's PIF. The other subcategory constitutes a range of non-resource-related performance improvement measures. The only example of these in Zambia was the ministry-level strategic planning system put in place under PSCAP, which aimed to create clear and consistent organization-level work plans that could be cascaded down to inform the individual-level performance management systems. Other examples in other countries included the system of organizational performance reviews and performance improvement plans created in Ghana under its Civil Service Performance Improvement Programme (CSPIP) or the organizational service delivery evaluations undertaken by Nigeria's SERVICOM for service delivery organizations in Nigeria. (For brevity, in the main text of the book, I will often refer to reform episodes without describing them in

TABLE 3.1 Content of civil service reforms in Zambia

	Public Service Reform Prog. (PSRP), 1993–1999	**Public Service Capacity Building Project (PSCAP), 2000–2005**	**Public Service Mgmt. Prog. (PSMP), 2006–2011**	**Public Service Transf. Prog. (PSTS) I & II, 2013–**
Individual performance mgmt.				
Leadership performance agreements		✓		✓
Individual staff appraisal & performance management	✓	✓	✓	✓
Org. mgmt. and capacity				
Org. reviews/ perf. improv. plans/ perf. mgmt.		✓		
Performance improvement funds		✓		
Service delivery-focused reforms				
Client-focused reforms			✓	
Sector-driven reforms				
Salaries and structures				
Staff redundancies & org. restructuring	✓	✓	✓	
Pay regrading/ reform	✓	✓	✓	✓
Other reforms				
Cross-government coordination				✓
Other				

Source: Author's synthesis based on document review and interviews.

full. In such instances, interested readers can refer to the relevant country narrative history in the appendix for more extensive details and evidence.)

The third category, service-delivery-focused reforms, captures two types of systemic reforms that aimed directly at improving service delivery (as opposed to upstream management issues of personnel and operations). The first subcategory, client-focused reforms, is exemplified by Zambia's effort to establish service charters that published information on services available, along with standards and timelines for accessing them, under its Public Service Management Programme (the successor project to PSCAP). Another example is Senegal's Comité d'allègement et de simplification des formalités et procédures administratives (CASPFA, Committee on Alleviating and Simplifying Rules and Administrative Procedures) in the late 1990s, which aimed to simplify procedures for the public to access services (as well as internal administrative processes). The second subcategory, which I dub sector-driven reforms, captures reform efforts that are systemic (in that they are centralized and span multiple sectors simultaneously) but have some inbuilt cross-sector variation in exactly what they aim to do. An example is Ghana's New Approach from 2009–2012, in which the secretary to the cabinet launched a coordinated push to get ministers to come up with sectoral strategies to improve service delivery in a number of priority sectors, particularly related to job creation and food security. As is evident from these examples, this category is relatively more internally diverse than the previous two categories.

The fourth category, reforms to salaries and structures, captures efforts both to downsize or "rightsize" staff numbers (as in Zambia's PSRP and PSCAP) as well as efforts to restructure pay scales and compensation systems (as also happened under PSRP and PSCAP). The former subcategory was often motivated by both fiscal savings and performance improvement, usually with some version of a "smaller but more efficient" logic, and can also include or be driven by organizational consolidation (as in Kenya's effort to reduce the number of government ministries under its Civil Service Reform Programmes and Economic Recovery Strategy in the 1990s and 2000s). The latter category was often driven by a desire to decompress salary scales—i.e., to increase the pay gap between junior and senior civil servants—as a way to indirectly incentivize performance by increasing the pay increments associated with promotion. Such reforms were also often intended to help retain more senior and more skilled staff, or to rationalize compensation by bringing off-salary benefits and allowances into core salary.

As its name indicates, the fifth category of reforms (other reforms) is a residual category for other types of systemic performance-oriented reforms that fit my definition of reform but did not fall into one of the above categories. One set of these is cross-government coordination reforms, which aimed not to improve the performance of organizations in isolation but to better coordinate their activities, particularly around key priorities. Examples of this include Ghana's repeated efforts to establish and reestablish this function with its Policy Coordination, Monitoring, and Evaluation Unit (PCMEU, 2001–2008), Policy Monitoring and Evaluation Unit (PMEU, 2009–2014), and Presidential Delivery Unit (2015–2016). A final residual subcategory captures all other reforms, ranging from Nigerian SERVICOM's creation of a weekly radio "help desk" program for service clients to air complaints and help resolve issues to South Africa's creation of a Center for Public Service Innovation to Senegal's creation of ministry-level specialists and units to coordinate training and personnel management in each organization (*conseillers en ressources humaines et organisation*, or human resource and organization councilors, and *cellules de gestion des ressources humaines*, human resource management units).

It's useful to get a sense of the relative frequency of each type of reform, but attempting to count and aggregate in this way requires many of the same caveats as the reform number calculations above. The boundaries between reforms (and thus the number of efforts) are unclear and subject to interpretation, and content categories and subcategories are each internally diverse and not necessarily mutually exclusive of other categories. While there's no perfect way to do it, the simplest and most transparent way to do so is simply to count the number of reforms or reform clusters that attempt each subcategory of reform, as per the six country-level reform content tables in the appendix.[19] The absolute number of these reform efforts could be counted in various ways, but our main interest here is in the relative frequency of each type, which is less sensitive to different approaches to classifying and aggregating reforms.

With those caveats again in mind, table 3.2 sums up the number of efforts to introduce each subcategory of reform across all six countries. The most frequent are individual performance management reforms (thirty-four instances), followed by salary and organizational structure reforms (thirty) and service delivery-focused reforms (twenty-seven). (Note that the aggregation of these reforms for the purpose of qualitatively representing content focus, taken from the reform content tables in the appendix, is different than the aggregation used for the rough estimate of reform frequency given above, so these figures on reform content aren't directly comparable.)

What's perhaps more interesting, though, is the frequency with which countries repeatedly attempt to implement the same type of reform over and over again. This pattern is most striking for individual-level performance management reforms: across my study period, and subject to the above caveats about counting and aggregation, these were attempted three times in Nigeria, four times in

TABLE 3.2 Content of reforms across countries

Type of Reform	Ghana	Kenya	Nigeria	Senegal	South Africa	Zambia	TOTAL
Individual performance mgmt.							
Leadership performance agreements	4	2	1	1	4	2	14
Individual staff appraisal & performance management	6	4	2	3	1	4	20
Org. mgmt. and capacity							
Org. reviews/ perf. improv. plans/ perf. mgmt.	1	3	3	2	3	1	13
Performance improvement funds	1	0	0	0	0	1	2
Service delivery-focused reforms							
Client-focused reforms	3	3	2	3	2	1	14
Sector-driven reforms	2	2	3	2	4	0	13
Salaries and structures							
Staff redundancies & org. restructuring	2	2	3	5	1	3	16
Pay regrading/ reform	3	2	2	1	2	4	14
Other reforms							
Cross-government coordination	3	2	2	1	1	1	10
Other	0	0	3	5	1	0	9

Source: Author's calculations.

Senegal, five times in South Africa, six times each in Kenya and Zambia, and ten times in Ghana. This pattern also appears with many other types of reforms, too—for example, Ghana attempted to introduce ministerial service charters as part of CSPIP (1994–2001), the Public Sector Reform Agenda (2006–2011), and the National Public Sector Reform Strategy (from 2016). These attempts to implement the same reform repeatedly hint at the less-than-full success of prior efforts to do so.

But while this is the most striking pattern in reform content, it is not the story of every reform. There are also many cases of governments trying to implement a reform once and then dropping it, as well as governments including a particular component in repeated reforms as a way to institutionalize and sustain it. Examples of the latter include the sustained use of Rapid Results Initiatives in Kenya across successive reform waves, Ghana's sustained use of Chief Directors Performance Agreements in the 2010s (in contrast to its failure to do so in the 1990s and 2000s), and Senegal's sustained progress in digitization of service delivery since the 1990s.

These patterns and variations in reform content naturally question how successful these reform efforts have been, to which we now turn.

How Successful Have Reforms Been?

The most striking observation that emerges from examining reforms' record of success is how little apparent variation there is: Not a single reform succeeded at achieving all its goals, but almost all achieved some of them. There is, of course, some variation across reforms in the degree to which each succeeded and in what ways, and I initially expected that this book would be focused on trying to measure and explain this variation. However, the most salient descriptive fact about reforms is not how large the gap between the most and least successful reforms is but how small it appears to be. Indeed, nearly every reform can be viewed both as a partial failure and as a partial success.

This pattern is difficult to show quantitatively because trying to formally code or quantify reform success is even more fraught than the (already heavily caveated) summary figures I presented above about reform frequency and content. Success for civil service reforms can only be judged against expectations, but these expectations might be too high or too low. This means that a very ambitious reform could achieve a lot but look like a failure, whereas a half-hearted reform with easy targets could get closer to achieving them despite changing little. In-depth qualitative research on reform efforts can often give a good (if sometimes contested) sense of both the ambition of the goals of the reform as well as its actual achievements of the reform, but trying to codify this into a dataset that could be summarized or analyzed quantitatively would do too much violence to this messy reality.

Instead of using quantitative summaries, though, we can illustrate that there is surprisingly little variation in achievement of goals by comparing two reforms: one that was perceived as relatively successful but still fell short of its goals, and another that was perceived as relatively unsuccessful but still achieved some improvements. An example of the former is Ghana's adoption of a Performance Improvement Facility (PIF) under the Civil Service Performance Improvement Programme (1994–2001), while an example of the latter is Zambia's adoption of a Performance Improvement Fund (PIF) under PSCAP (2000–2005) that I described above. The design of the interventions in each case was very similar: the creation of a discretionary fund administered by a central reform institution and financed by a donor grant, to which individual ministries could submit proposals for rapid disbursement of small amounts of money to meet the costs of innovative, demand-driven service delivery improvement initiatives, linked to a larger process of strategic organizational planning and performance review.

But whereas Zambia's PSCAP was widely seen as donor-driven, generated jokes about how it paid mostly for cars, and was terminated early by its donor partner, Ghana's CSPIP was almost universally seen as "homegrown" and driven from within the government itself by a charismatic head of civil service.[20] It established a deeply participatory and thorough performance review and improvement planning process that linked closely to the PIF, was generally seen to be generating positive impacts, and its donor partner wanted to extend it for a second phase.[21] Ghana's PIF under CSPIP was carefully administered by a highly motivated team that undertook extensive scrutiny and monitoring of each application and disbursement.[22] Internal records of the first years of Ghana's PIF showed that the disbursement committee not only applied a high quality filter to applications (funding only one of the initial five received) but also sought to remedy these perceived problems by undertaking a set of workshops for ministries to help them improve their conceptualization of and applications for these funds.[23] Ministries took this seriously and reportedly began competing with one another to access the PIF.[24] Nana Agyekum-Dwamena, who served as secretary of the PIF's disbursement committee as a young civil servant during the late 1990s, recalled applications from local governments for containers to put rubbish in as part of sanitation drives and giving the Passport Office funds for a generator so it could work through power outages. While the CSPIP reform was curtailed by changes in political and bureaucratic leadership after 2000, during its years in full swing, it represented something close to an ideal scenario for serious, innovative, government-driven civil service reform.

But Ghana's PIF still fell short of its goals. One review found that "although many organizations duly prepared their performance improvement plans and were rewarded with small grants, in a larger sense little overall improvement in performance resulted."[25] Agyekum-Dwamena also highlighted the mismatch between having funds available to support about ten small projects per year through the PIF and having about seventy organizational performance improvement plans waiting to be implemented at the height of CSPIP.[26] Another officer said that the project's five-year review found that, while the project was delivering its anticipated outputs, it was failing to achieve the outcomes envisioned for the PIF.[27]

At the same time, Zambia's PIF under PSCAP was far from a total failure. The program was heavily criticized for funding mainly asset purchases, but as one of its administrators explained, there were also real needs for computers, cars, and other logistical tools in many cases.[28] It also paid for some innovative projects that won international recognition and were sustained beyond the project's lifespan. And while PSCAP's internal review criticized the PIF because only fifteen of thirty-eight projects funded under the PIF had "an observable impact on service delivery,"[29] one could also take the glass-half-full perspective that the PIF managed to have observable impacts in fifteen different areas of service delivery. Though it fell far short of its expectations, it wasn't as if nothing was achieved by Zambia's PIF.

These two patterns were not restricted to PIFs as a reform category. Rather, they repeated themselves across the other types of reforms I studied: (1) reforms didn't all fail, but they did all fall short of fully achieving their goals; and (2) some reforms were more successful than others, but these distinctions were narrower and more ambiguous than one might expect. These patterns were also consistent across countries—it's simply not the case that reforms were successful in some countries and unsuccessful in others. One veteran of many reform efforts reflected that reforms have been about 60 percent successful,[30] for example, while another commented that "the reforms were not a total failure, but they did not achieve the transformation that was expected."[31] (For those who are interested, the country-by-country reform histories in the appendix give more details on the successes and shortcomings of each reform episode and reform era in each country.)

This pattern of reforms falling short of their goals is not restricted to Africa. Indeed, much of the literature on civil service reform in OECD countries has remarked on similar patterns in these contexts.[32] A classic article titled "Why Civil Service Reforms Fail" by Charles Polidano begins: "Most reforms in government fail. They do not fail because, once implemented, they yield unsatisfactory outcomes. They fail because they never get past the implementation stage at all. They are blocked outright or put into effect only in tokenistic, half-hearted fashion."[33]

I don't mean to imply that the patterns or causes of reform outcomes are the same everywhere. This literature should instead act as a caution to the common tendency outside (and sometimes even inside) African governments to assume that disappointing reform results must be due to specific challenges of administration or politics in Africa. This assumption can lead us to seek out explanations that mainly pertain to the context rather than to the difficulty of changing performance in large bureaucratic systems more generally. I won't attempt to explain here why we observe these patterns, as that is (part of) the job of the next two chapters. But from a purely descriptive standpoint, the story of reform success is clear: Every reform fell short of its goals, but reforms were not all complete failures, and most reforms did achieve something.

DRIVERS OF REFORM

The final question this chapter addresses is: What actors and factors shaped the timing and design of reforms? I consider the evidence for three potential explanations: politics, donors, and ideas.

Politics

One set of theories about what drives civil service reform centers around the role of national political leaders and dynamics. For example, scholars like Mai Hassan have shown how political leaders manipulate the posting of senior bureaucrats for their own political advantage, and Sylvester Obong'o has examined how resistance by politicians to reforms that would reduce their patronage powers—such as redundancies and greater professionalization of hiring and promotion powers—explains much of Kenya's history of failing to implement reforms.[34] Work by Daniel Appiah and Abdul-Gafaru Abdulai and by Frank Kwaku Ohemeng and Felix K. Anebo also highlight the role of electoral cycles that lead to leadership turnover and policy discontinuities as a factor in disrupting the continuity of reforms.[35]

Within my reform data, I found many examples of reforms that were interrupted or otherwise went unimplemented due to these dynamics but also many counterexamples and patterns that are not well explained by these theories. The picture that emerges is one in which these and other political factors play important roles but nonetheless are not fully determinative of the timing, content, and implementation patterns of reforms.

With respect to the idea that politicians might resist fully implementing politically painful reforms, there are, of course, numerous examples of this both in secondary evidence and from my interviews. However, these examples were heavily concentrated within a single category of reforms: those that related to downsizing employee numbers and/or reducing the number of ministries, such as Zambia's PSRP, Kenya's CSRP, or Nigeria's Oronsaye Committee on Rationalizing the Structure of the Federal Government. This makes sense, given that these reforms directly threaten the livelihoods of urban elites and middle classes, some of whom may be political supporters, and threaten to restrict the number of patronage appointments politicians can dole out both at the highest level (e.g., ministers) and at the frontlines (e.g., teachers, drivers, cleaners).

But the reasons why politicians might object to the other categories of reforms, such as the use of organizational performance reviews, one-stop service centers for clients, or policy coordination units, are less obvious. If anything, political leaders generally stand to gain from improvements in service delivery, and there are numerous examples of senior political leaders and their advisors being strong and proactive advocates of these types of reforms. Examples of such full-throated support at the highest political levels for service delivery-oriented civil service reforms include the Economic Reform Strategy and Results for Kenya initiatives from 2003–2007 under President Kibaki, South Africa's Batho Pele initiative from 1997 under President Mandela, and Ghana's Public Sector Reform Agenda from 2006 under President Kufuor. Self-interested political opposition certainly helps explain the failings of some types of reforms but is not a universal explanation for them.

There is also a second category of reforms that sometimes fall victim to opposition from politicians but for which the politics are more nuanced: individual performance-linked incentives. As I discuss in more depth in the next chapter, senior political leaders typically support these systems in the abstract, and there are even examples of presidents choosing to introduce them for their ministers in Nigeria, Senegal, and South Africa. However, the implementation of these systems generates costs both for politicians and managers that often lead to their abandonment or nonimplementation. So while they are also a type of reform for which political considerations turn out to be very relevant, their politics can't be boiled down to a matter of crude support or opposition from politicians.

What of the role of leadership transitions that result in the discontinuation of reforms initiated under the previous leader? There are certainly some examples of reforms that correspond to political cycles in this way: the high-profile Economic Reform Strategy (2003–2007) that started under President Kibaki's administration in Kenya, President Obasanjo's launching of the Public Service Reform Strategy in 2003 shortly after winning a second term in office, President Zuma's adoption of Ministerial Performance Agreements on taking office in 2009, and so on. But there are also many examples of reforms that span changes in political administrations, such as Ghana's Single Spine Pay Policy reform (initiated under President Kufuor, implemented under President Atta-Mills), Kenya's continued use of Rapid Results Initiatives and Performance Contracts (initiated by President Kibaki, continued throughout the Grand Coalition Government with Prime Minister Odinga), or Nigeria's continued effort to reform its Annual Performance Evaluation Report system (spanning the administrations of Presidents Yar'Adua, Jonathan, and Buhari). In other cases, a reform appears to have been discontinued and a new one initiated, but the actual content was maintained. For instance, President Wade's administration in Senegal had developed the Schéma Directeur de la Réforme de l'Etat (SDRE, State Reform Master Plan) in 2011 but lost elections in 2012. Under President Sall, it was relaunched in 2013 as the Schéma Directeur de Modernisation de l'Administration Publique (SDMAP, Public Administration Modernization Master Plan), with verbatim identical components and summary diagram.[36] So while there is some evidence to support the idea that leadership turnover can undermine reform progress, there are also plenty of counterexamples.

The example of Senegal's retaining the content of the SDRE under the SDMAP—despite the SDRE having been designed during the administration of a political opponent—illustrates another important observation about the politics of systemic civil service reforms: Many reforms deal with relatively dry and obscure administrative rules and processes that do not correspond to existing political cleavages and inspire little political mobilization. Mass layoffs are highly politicized, but the finer points of organizational performance reviews or internal structures of service delivery agencies tend to be lost not just on voters but also on most politicians. A large proportion of the existing research on the politics of reform focuses on highly politicized types of reform precisely because they are the most visible and, in some ways, are the most attractive to study, but examining the broader universe of reform initiatives pursued by countries also

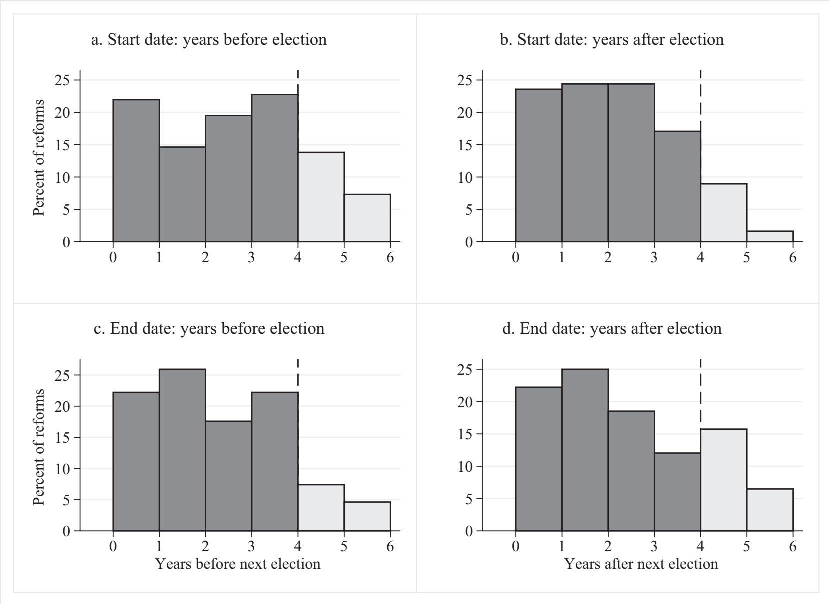

FIGURE 3.3 Comparing Reform Cycles to Electoral Cycles

Source: Author's calculations.

reveals many reforms for which political factors were less salient or were focused more on the relatively mundane workplace politics that exist in any large organization anywhere in the world.

Can we detect any general patterns with respect to how reform cycles do (or do not) coincide with electoral cycles? One way to do this is to compare the timing of the start and end of reforms with the dates of elections in each country. Figure 3.3 does this by showing (a) how many years before the next election each reform started; (b) how many years after the last election each reform started; (c) how many years before the next election each reform ended; and (d) how many years after the last election each reform started.[37] This has some methodological challenges and limitations as a test of the impact of political cycles on reforms, but if political cycles are a dominant factor in determining reform cycles, we should see at least some visual evidence of it in these graphs.

There is no obvious pattern of reform timing with respect to elections on any of these graphs. (The lower numbers of reforms after year four of each graph are driven by the fact that most gaps between elections are not that long due to shorter electoral cycles and deaths in office, so those are drawn in a lighter shade.)

Although there are certainly cases of reforms being initiated or terminated due to leadership transitions and political cycles, in the aggregate, we see little evidence that political cycles are the main factor driving reforms.[38]

This is not to say that political factors don't play an important role in driving some aspects of the design and implementation of civil service reforms. They obviously do, and the reform histories in the appendix highlight numerous instances when reforms were driven or undermined by political imperatives—indeed, political factors often impelled reform while simultaneously undercutting its implementation. But political incentives on their own are only part of the explanation for the broader patterns we observe.

Donors

A great deal of research has explored how reforms in Africa (and elsewhere in the Global South) have been driven by the incentives and dynamics of international donor institutions and their interactions with national governments. To highlight just a few examples from this extensive literature: Nicholas van de Walle describes how African governments in the 1980s and 1990s strategically avoided implementing reforms that international financial institutions demanded of them in order to access loans and grants as part of structural adjustment programs;[39] Matt Andrews writes of international donors ignoring local context by naïvely transplanting "best practice" reform models and of developing countries "adopting reforms as signals" in order "to garner short-term support from the international community";[40] Rosina Foli and Frank Kwaku Ohemeng argue that reforms often fail because international bureaucrats in donor institutions do not adequately understand local contexts and operate according to different logics than recipient governments do;[41] Mark Buntaine and colleagues find that donors' institutional reform projects often choose "shallow" targets related to what the reformed institutions should look like rather than targets related to effectiveness or service delivery outcomes because these are more attractive both to donor agencies and country governments;[42] Tunji Olaopa attributes reforms' focus on measurable short-term gains over more meaningful but longer-term behavioral changes to donors' need to "show quick results to convince their domestic constituents";[43] and many studies have examined the role that international institutions play in diffusing norms, ideas, and reform fads around the world.[44]

These dynamics are all evident across the reforms in my data to varying extents. But both from documentary records and from interviews, it is equally clear that the activities and incentives of international donor agencies are only one factor among many in the design, implementation, and outcomes of these reforms.

One way to examine the limitations of donor influence is simply to count how many of the reforms I identified corresponded exactly to projects run by international donors. Of the 131 reforms in all six countries (using the counting methodology described above, with all its caveats), just thirty-one—under a quarter—were identified as projects by a donor organization like the World Bank, UK DFID, or UNDP. Some of these were generic omnibus loans or grants that included a civil service reform component (e.g., Senegal's Structural Adjustment Program IV, 1990–1992), others were donor projects that corresponded directly to a government reform policy in that they had the same or similar name, dates, and content (e.g., Kenya Civil Service Reform Programme II, 1998–2001), and others were standalone donor projects that funded government reform activities but had no directly analogous reform policy or program in government (e.g., DFID's Federal Public Administration Reform project in Nigeria, 2011–2016).

Of the remaining ninety-nine reforms, many—it's impossible to definitively say what fraction—had some financial support or technical assistance from donors. But many did not or only intermittently had it.[45] And while some of these reforms may have been designed in the shadow of donor influence or with an eye toward trying to attract funding or signal a commitment to reform, at the end of the day, these ninety-nine reforms were neither designed nor implemented by donor institutions. Reformers' choices may have been constrained by donor pressure or willingness to provide financial support, but governments were still the ones navigating these constraints and deciding whether and how to invite donor involvement.

Even among reforms that corresponded directly to donor funding projects, governments often exerted more agency than is typically assumed. At one end of this spectrum lie reforms like Ghana's Civil Service Performance Improvement Programme (CSPIP, 1994–2001), which was almost fully funded by UK DFID but was wholly designed and driven by reformers in the Office of the Head of Civil Service (OHCS). Indeed, OHCS had approached DFID after rejecting a new project offer from the World Bank that wanted OHCS to continue the cost-cutting and downsizing focus of the preceding years. This "homegrown" approach to reform within the framework of a donor-financed project was facilitated by a strong head of civil service with a clear reform vision as well as progressive thinking by DFID staff "who bought into our methodology" and gave OHCS a "free hand" to formulate the reform.[46]

At the other end of the spectrum lie reforms like Kenya's Civil Service Reform Programme (CSRP, 1993–1997 and 1998–2001) and Zambia's Public Service

Reform Programme (PSRP, 1993–1999). These reforms were (and mostly still are) widely perceived to have been designed and driven by donors and bore many of the hallmarks of donor-driven reforms: direct correspondence of the reform name to a donor project, majority donor funding of activities, donor pressures for downsizing of staff and regrading of salary scales to increase inequality between the highest- and lowest-paid officers, and the adoption of internationally legitimated reform instruments like performance-linked incentives.

But interviews with some of the civil servants who were involved in the design and implementation of these reforms reveal a more nuanced picture. One civil servant who had worked in the Zambian Management Development Division, which was the lead institution within government for the PSRP, emphasized that within his team, there was both a recognized need for and commitment to change and that they insisted on government staff being at the forefront of both design and implementation (despite occasional involvement from consultants).[47] In his view, the donor conditionalities that were in place were a kind of backstop to ensure that the reforms were implemented, but he and his colleagues felt like they owned the reforms.

Similarly, most observers perceived Kenya's CSRP and the Kenya Civil Service Reform Programme and Action Plan that laid the foundations for the CSRP as donor-driven, and the CSRP itself was a largely donor-funded program.[48] But while fiscal crisis and donor dependence did drive President Moi to take actions (such as compulsory redundancies and reduction of the number of ministries) that were half-heartedly implemented at best, many senior public servants were also concerned about the country's trajectory and saw a need for reform.[49] When asked whether the CSRP was donor-driven, for instance, one former civil servant closely involved in implementing the CSRP replied:

> No, the Civil Service Reform Programme and the Action Plan was actually put together by civil servants. . . . We did have a few people we would share our thoughts with. One of them of course was at that time the Swedish government, there was an officer who would come and meet with us quite regularly. I think it was only that person, because the other times we would then meet maybe after a quarter just for the UN basket[-fund] people to find out how we're progressing and how we're using available funds. But they did not direct where those funds should go.[50]

When I asked another civil servant involved in several reforms in Ghana in the 1990s and 2000s whether they were donor-driven, he first replied, "Reforms are almost always donor-driven," but then proceeded to explain that there was agreement across the board that reforms were needed and that donors didn't bring their own agenda and tell countries to "take it," with the details of reforms instead being a product of mutual discussion."[51]

The idea that African governments exert agency in their negotiations with international donors is a theme of the literature on aid more generally. For example, Folashadé Soulé-Kohndou shows how bureaucrats in Benin find ways to overcome asymmetrical power relations (both with respect to Chinese donors and their own political leaders) in order to improve infrastructure project agreements.[52] Similarly, the contributors in Lindsay Whitfield's edited volume *The Politics of Aid: African Strategies for Dealing with Donors* emphasize the ways that African governments are far more than passive recipients of donor programs and contrast their approach to the donor-centric assumptions of much of the literature.[53]

In his research, Sylvester Obong'o presents yet another argument against centering the role of international donors in the analysis of reform implementation. He points out that Kenya—like many other African countries—launched several reform commissions in the two decades after independence that were endogenously driven and designed specifically for the Kenyan context, with little influence from international donors. However, the implementation record of these reform efforts was no better than for the more donor-influenced CSRP in the structural adjustment era of the 1990s. The poor implementation record of these reforms cannot, therefore, be explained simply by the presence or absence of donors.[54] Similarly, the presence of significant reform implementation issues in OECD countries (as noted above) as well as in countries that relied relatively less on donor funding (e.g., South Africa and Nigeria) suggests that while the dynamics of donor involvement do sometimes play important roles in reforms, they are far from the sole cause of implementation challenges.

One reason why a large share of research on civil service reform in Africa (and the Global South more broadly) focuses on international donors is likely that these reforms are usually more visible and easier to study than reforms where donors play less of a role. Donor institutions produce huge volumes of documents, reports, and project success metrics that are often made publicly available; donor staff are often easier for researchers to access than government officials; and the bureaucratic and cultural logics of donor institutions mean that they are often systematically trying to compare performance and learn lessons across many projects and countries, so they produce the kinds of data and processes that facilitate researchers' efforts. In conducting research for this book, my research assistants and I were able to identify nearly every reform in which donors were heavily involved from our initial review of existing donor project databases and secondary literature. In contrast, we only identified many government-initiated reforms in the course of interviews with civil servants because there was usually much less written documentation on them in the public domain. Studying reform by focusing only on donors or using only donor-produced data would have produced a partial, skewed picture of the reality of reform in each country.

A second possible reason why attention tends to focus on donors' role in driving reforms is that the normative and epistemological frameworks through which many observers view governments in Africa make it easier to attribute agency to international actors who are perceived to be intervening from "outside" the system, rather than to actors who are operating "within" it. Normatively, it is easy for many people (again, especially those based outside Africa) to see donors or NGOs as imperfect but well-intentioned actors altruistically trying to help solve problems, while African politicians or bureaucrats are seen as self-interested principals or agents embedded in a negative equilibrium they can't escape from. Methodologically, assuming that international donors are outside the system they're intervening into makes it possible to treat their interventions as exogenous (in certain cases at least) and thus able to be evaluated using social scientists' causal inference toolkit in a way that would be harder to justify for endogenously designed reforms. This is not to criticize researchers that do focus on the role of international donors—they are important actors, and studying them from various angles is important—but rather to advocate for the importance of also recognizing the agency and crucial role of reformers within government.

Ideas

What else, then, drives the actions of individuals and organizations working within these governments with respect to the design and implementation of civil service reforms? As in any organizational field, there are, of course, many possible factors: public service motivation, prestige, the search for legitimation from other members of their profession, desire to do their job well, and pecuniary self-interest. And all these played important roles to different extents for different people at different moments.

In the remainder of this chapter, though, I want to focus on one set of factors that emerged as especially powerful and pervasive in my interviews: reformers' own ideas, understandings, and mental models about the problems that reforms were trying to address and how they might do so. Civil servants were not just inert, passive actors responding to external pressures. They used their experience, training, and available evidence to come to understandings about what wasn't working in their institutions and why, and they formulated hypotheses and plans about how to fix it. These ideas and mental models thus guided the choices they made about what issues to prioritize, what reform instruments to adopt and try to implement, and how to navigate the external pressures, constraints, and opportunities within which they were working.

I went into this research expecting to hear stories from civil servants about how they were forced by donors or politicians to adopt reforms that they didn't believe in. These complaints did sometimes arise, but they turned out to be relatively infrequent. Instead, interviewees overwhelmingly highlighted the failure to implement reforms as the problem rather than the objectives or content of the reform per se. For instance, one rank-and-file civil servant in Ghana stated, "The objectives of [the] reform were good, but during the implementation of the reform, there is always an issue."[55] Similarly, a former civil servant who was instrumental to many of South Africa's reforms lamented, "The sad part was that [the reform] was all very well on paper, it just never translated into practice."[56] There were, as with public servants everywhere in the world, frequent laments about perceived political barriers to reform success, such as lack of political will or leadership turnover. But most of the reform ideas and directions themselves (with the partial exception of staff cuts or downsizing efforts linked to structural adjustment programs in the 1980s and 1990s) were broadly perceived by most interviewees as positive ones for the civil service—if only they could be implemented.

There were also numerous examples of reform leaders and ordinary civil servants speaking out in favor of reforms that had been adopted in ways that went above and beyond the standard level of support for government policies that public officials are expected to show in routine press conferences or meetings with donors. This was most evident through the direct engagement of reform leaders with academic research. To cite just a few of many examples, Ghana's then-Head of Civil Service, Robert Dodoo, authored a 1997 article in the academic journal *Public Administration and Development* on how the CSPIP reform was taking on the task of setting and measuring performance standards for organizations and institutions.[57] In 2006, Margaret Kobia (then Director of the Kenya Institute of Administration, later Chairperson of the Public Service Commission and Cabinet Secretary for the Ministry of Public Service, Youth and Gender Affairs) and Nura Mohammed presented a paper to the African

Association for Public Administration and Management on Kenya's experience with performance contracting, giving a detailed international history of the idea, how it came to be adopted in Kenya, and the successes and challenges of its implementation, based in part on survey research they had undertaken with Kenyan civil servants.[58] And Geraldine Fraser-Moleketi, South Africa's Minister of Public Service and Administration from 1999–2008, wrote a detailed and thoughtful master's thesis in 2006 documenting and analyzing South Africa's postapartheid reform journey.[59]

A number of interviewees expressed concern about uncritical transplantation of reform ideas from other countries into African contexts without appropriate adaptations. However, most civil servants I interviewed mainly expressed this as a criticism of "us" (i.e., civil servants) rather than "them" (i.e., donors). At the same time, many reformers in government were actively seeking out experience from the reform trajectories of other countries in Africa, elsewhere in the Global South, and in high-income countries. The process of formulating Kenya's dynamic and internally driven Results for Kenya program (2004–2008) included study visits to the UK, Sweden, and Canada, whose Results for Canadians program directly inspired the Results for Kenya name.[60] South Africa's Management Performance Assessment Tool (2011–2016) was inspired by Canada's Management Accountability Framework, which had been encountered during a study visit.[61] And several reforms in Zambia were modeled on or benchmarked against similar reforms in Ghana.[62] This view was expressed neatly by Robertson Nii Akwei Allotey—then acting Chief Director of Ghana's Ministry of Public Sector Reform, subsequently a Commissioner of the Public Service Commission—in a 2008 interview with Princeton University's Innovations for Successful Societies program:

> I would say that the—one cannot be an island, especially Ghana, cannot be an island at all. . . . Citizens' charters, for example, are good for everybody no matter where they are, Singapore, America, etc., because then it puts the responsibility on the organizations to perform, and Ghana has borrowed them, and we are working with that . . . So, it's nice to borrow, but at the same time, you adapt it to suit your circumstances, and that's what we are doing in Ghana here. We are also lucky, because other African countries come to Ghana to learn exactly what we are doing here and try to adapt it to what they have in their various countries. I've been very fortunate to have participated in conferences abroad, and I speak to the issues concerned, and they also are very keen to learn as to what we are doing here. So, it's more of sharing best practices and experiences. We are so welcome to that, because it helps us in a way. [63]

While adaptation of reform ideas to local contexts is always both challenging and important, the global diffusion of reform ideas is driven not just by uncritical mimicry or the imposition of "best practices" by donors as is sometimes assumed but also by a real desire on the part of civil servants to gain inspiration and learn lessons from other countries' experiences. The myriad ways that civil servants processed their experience and training, formulated plans and advocated for them, and consciously and critically sought to learn from other countries' experiences illustrate how important civil servants' own ideas about performance and reform were for the choices they made.

Having said that, I don't mean to suggest that the ideas of key reform actors in government were the sole drivers of reforms. Political incentives and constraints, as well as pressures from international donors, certainly helped shape the trajectory of reforms in each of the six countries I studied. And—at the risk of stating the obvious—civil servants often disagree among themselves about everything from how future reforms should be approached to the origins of past reforms.

But the reasons that I nevertheless emphasize the role of reformers' own ideas and agency are threefold. First, this theme emerged from my interviews far more strongly than would be indicated by much of the existing literature and theory. Second, failing to note this would do a disservice to the thought and dedication that so many public servants in these six countries and around the world have poured into reform efforts over the years. And third, if ideas and mental models do matter—even a little bit—then there is at least some scope for more evidence and new theory to help current and future generations of reformers better understand their institutions and formulate strategies for improving them. This is a hopeful perspective for reformers and researchers alike.

• • •

This chapter has aimed to distill the empirical richness of the detailed narrative histories in the appendix into a number of big-picture descriptive findings and trends. In doing so, it has covered a lot of ground, so let's briefly summarize the four key messages. First, all six countries undertook many reforms, and these reforms overlapped one another, so the picture was not one of occasional spurts of reform but of constant and interrelated reform efforts. Second, many of these reforms entailed repeated efforts to implement the same type of reform. Third, these reforms have universally fallen short of achieving their full expectations, but they were not all failures, and there were many meaningful successes. Fourth, political factors and donor involvement both contributed to shaping the adoption, design, implementation, and outcomes of reform, but so did the ideas, mental models, and agency of reformers within government.

The next two chapters turn to the task of describing two common mechanisms of reform—focusing on formal rules and structures (chapter 4) and approaching reform as a one-off intervention (chapter 5)—and analyzing how they led to the observed patterns of implementation and impact that this chapter has summarized. Both mechanisms were driven in part by structural factors but also in part because they corresponded to mental models of how bureaucracies work and how to improve them that are commonly held by civil servants, politicians, donors, and scholars alike. Understanding these two *mechanisms of failure* is a key part of explaining the track record of past reforms and of beginning to envision an alternative approach.

4
The "What" of Reform

Focusing on the Formal, Neglecting the Informal

If you want civil servants to perform their duties better, why not reward people who perform well and punish people who perform badly? In the messy and complicated world of government bureaucracies, this intuitive and apparently simple proposition has motivated countless reforms around the world over the past several decades.

Indeed, the introduction of policies that attempted to link individual performance to some type of reward or sanction was the most common type of reform in the six countries studied in this book. There were thirty-four efforts to establish and/or operationalize such systems, both for rank-and-file civil servants as well as for senior-level leaders—just over a quarter of all reforms. These policies have also been popular elsewhere around the world.[1] In the Organisation for Economic Cooperation and Development (OECD) group of high-income countries, for instance, thirty-two of thirty-seven member states had some version of these policies in place for their senior managers as of 2019.[2]

But while individual-level performance-linked incentive policies hold an obvious appeal, among the reform efforts I studied, their track record of implementation was dismal. Of the thirty-four performance-linked incentive reform efforts in these countries, *zero* succeeded in sustainably delivering differentiated rewards and punishments. Only two delivered differentiated financial rewards at all—both of which ceased doing so within a few years—but neither of these delivered sanctions for poor performance. In the majority of such reform efforts, systems quickly converged to an equilibrium where nearly every employee received the same score, and incentives (if they were given at all) were not differentiated according to performance. Indeed, the fact that countries repeatedly tried to introduce such schemes is evidence of the shortcomings of preceding efforts to do so.

Viewed in isolation, the failure of each particular effort might be attributed to failings in implementation, cultural mismatch, a lack of political will, or some other combination of idiosyncratic factors. For instance, a 2008 review of Zambia's Annual Performance Appraisal System (APAS, part of the PSRP and PSCAP reforms discussed in the previous chapter) argued: "Users of APAS rarely blame its limitations on the design of the system itself. More often, the blame is placed on the human factor, which includes lack of management support, inadequate resources and little commitment to its implementation. . . . In summary, the basic problems with the current system are in its implementation and, as explained earlier, the lack of a supporting performance culture."[3]

But viewed together, the results of these thirty-four reform efforts appear not as isolated implementation failures but as a *repeated pattern* of nonimplementation, despite a widespread belief that they would be an important—even necessary—lever for improvement. I argue that this pattern emerged because of an inherent mismatch between the highly formalized rules and processes introduced to administer the incentives and the often unformalizable sets of actions that civil servants must undertake to perform their duties effectively. Faced with this mismatch, one of two things happened. Most frequently, the targets set for individuals were vague and/or obviously incomplete, which made it difficult to objectively rate each individual's performance, and so concerns around fairness and morale meant that almost everyone ended up with the same score—which undermined the whole point of the incentive system. In the instances when reformers insisted on imposing objectively measurable targets on individuals and linking their measured achievement to high-powered incentives, this quickly led to civil servants distorting their efforts to meet targeted goals while ignoring nontargeted ones and/or finding ways to water down their targets. In the face of complaints and resistance from actors inside and outside the bureaucracy, such systems were either abandoned or became much less rigorously implemented. Either way, the differentiated incentives that were meant to act as carrots and sticks to spur better performance always failed to be systematically and sustainably implemented.

This chapter explores how focusing on the formal so often became a mechanism of failure for reforms. To do so, the first two-thirds of the chapter zooms in on individual-level performance-linked incentive reforms in these six countries, as they provide perhaps the clearest illustration of this mechanism. But focusing on the formal as a mechanism of failure wasn't unique to these six countries, nor was it restricted to individual-level performance-linked incentives. The final third of the chapter, therefore, zooms back out to argue that such patterns also exist with performance-linked incentive policies in many other countries around the world as well as with other types of reforms within these six countries.

REPEATED EFFORTS TO INTRODUCE INDIVIDUAL-LEVEL PERFORMANCE-LINKED INCENTIVES

At the beginning of this book's study period in the 1980s and 1990s, all six countries had legacy staff appraisal systems that essentially consisted of each officer's supervisor filing an administratively oriented and confidential annual report about them. These typically involved no participation, target-setting, or formal assessment of performance against targets. In 1991, for instance, Ghana's then-head of state, Jerry Rawlings, lamented that the "confidential reports on individual performance are just a matter of routine; almost everybody, that is, the hard-working and the lazy, get a good confidential report."[4] Furthermore, the results were often withheld even from the officers themselves. As one Zambian civil servant remarked of these old-fashioned legacy systems, "That Annual Confidential Report was so confidential you wouldn't even know what is in it!"[5] Where they were shared, as in Kenya, the assessments were implemented perfunctorily, with the outcome being conveyed in an impersonal letter that merely conveyed the absence of any adverse findings.[6] There was thus a widely shared perception that making these annual appraisal systems more performance-oriented and linking them to rewards and/or sanctions should be at the top of the reform agenda.

In Ghana, Kenya, Senegal, South Africa, and Zambia, efforts to introduce formal performance assessments and incentives formed crucial parts of the first generation of civil service reforms in the 1980s and 1990s. (Nigeria also began to undertake a similar reform but slightly later.) Staff appraisals were to be participatory, involving joint target-setting between officials and their supervisors at the end of the year combined with formal assessment and feedback at the end of each year. This assessment was intended to be linked to pecuniary incentives like promotion decisions, financial incentives, and/or potential job termination. Broadly inspired by ideas associated with the New Public Management movement and similar efforts around the world, in each country these reforms were first implemented in the context of downsizing and structural reforms that aimed to create smaller but higher-skilled, higher-paid, and hopefully more motivated civil services.

In each country except Nigeria and South Africa, these systems were adopted as part of donor-linked programs, but they also had broad support from governments, which envisioned them as ways to reward performance and improve remuneration. In Ghana, for instance, one internal planning document (undated, but from around 1991) stated, "When it is known by the members of the organization that decisions on promotion, salary increment, training and dismissals shall be fairly made on the basis of performance appraisal results, then the exercise achieves high respectability and serves as a motivator."[7] An internal circular in the same year directed ministries to set aside 10 percent of their personnel budgets for the provision of merit-linked cash and noncash awards.[8] In 1993, minutes of a meeting on the CSRP stated, "It should be made easy to remove non-performers from key positions."[9] Similarly, in Zambia, a 2002 report that laid the groundwork for PSCAP's pay and incentive reforms stated: "The newly articulated pay policy should as much as possible, aim at explicitly linking pay to performance, signalling a major change in the incentive system and in performance expectations. Rewards and penalties are both vital for a well functioning incentive regime. . . . Meaningful performance incentives are a must."[10]

In each country (again, except Nigeria), the new appraisal systems were formally established within a few years. However, the implementation of these systems fell far short of expectations. First, no country established any rewards or punishments for good performance, as the illustrative quotes in table 4.1 highlight. Second, to varying extents across countries, years, and organizations within each country, many civil servants did not even undertake these assessments each year—the formal system was simply not consistently enforced, and the lack of consequences for not doing the appraisal led to either the supervisor or supervisee neglecting to complete the process.[11] Recall, for example, the practice in some ministries in Ghana (described in chapter 2) of officers completing multiple years' worth of appraisal forms with different dates at the same time and getting them signed by their supervisors in order to be eligible for promotion—a practice which was also reported by rank-and-file interviewees in Zambia.

The disappointing results of this first wave of incentive-linked performance appraisal reforms led in each country to repeated efforts over the following decades to reintroduce or effectively operationalize similar systems—except in South Africa, where such efforts were focused more on senior managers than rank-and-file staff (see below). These sometimes modified the formal details of the system—the name of the system, format of the appraisal template, stipulated timing or process for supervisor-supervisee discussion, type of incentives to be offered—and other times were simply announcements of renewed efforts to get officers to actually complete the appraisal process or of the intention to link incentives to measured performance.

Table 4.2 illustrates the number of times such reforms were attempted in each country. Ghana, for instance, launched its first performance-linked appraisal system in 1991. In 1993, the minutes of an internal governmental meeting record the then-head of civil service as stating, "The merit pay system was well designed but it is not working. The government has a lot to gain by solving problems

TABLE 4.1 Illustrative quotes on the failure of individual performance-linked incentive systems

Ghana

- "I cannot rely on this instrument to tell me anything. . . .Everybody is very very good, but you and I know that when it comes down to productivity, not everybody is excellent."

Kenya

- "[It] was being taken as a routine thing . . . even if your performance was not very good, nothing would happen to you. You would still be getting your salary, you still even get promoted, and so on. So it wasn't really taken very seriously. . . ."

Nigeria

- "APER is not useful, it does not assess anyone."

Senegal

- "We evaluate in a routine, mechanical way. One does the evaluation, gets a rating, and gets promoted. . . . But in reality, we haven't sufficiently integrated the dimension of officers' performance to improve the quality of services."

South Africa

- "I don't think there was a single public manager dismissed from the public service because of poor performance . . . lots of these things were put in with good intentions, they were simply just watered down to an extent that they just became tick box exercises . . . everyone signed the agreement, everyone did the assessment after six months, everyone did the annual assessment, and if you look at the most of those assessments, everyone got their average assessment, so they got their performance [increment] on an annual basis and they were quite happy with that."

Zambia

- "APAS has mainly been used for administrative convenience. . . . I have never seen someone be demoted due to bad performance."

Sources: Quotes in alphabetical order of country: interview, GHA13; interview, KEN2; interview, Tunji Olaopa; interview, Ibrahima Ndiaye; interview, SA5; interview, ZAM17.

TABLE 4.2 Reforms aiming to introduce or operationalize individual performance-linked incentives

	Annual performance appraisals (for rank-and-file staff)	Performance contracts (for senior leadership)
Ghana	• Civil Service Reform Programme, 1987–1993 • Civil Service Performance Improvement Programme, 1994–2001 • Public Sector Reform Agenda & Single Spine Pay Policy, 2006–2011 • Performance Management Policy, 2007–2009 • Performance Management Policy for the Public Services of Ghana, 2012– • National Public Sector Reform Strategy, 2016–	• Performance Agreement System, 1996–2000 • Public Sector Reform Agenda & Single Spine Pay Policy, 2006–2011 • Minister Performance Contracts, 2013 • Chief Director Performance Agreements, 2013– • National Public Sector Reform Strategy, 2016–
Kenya	• Civil Service Reform Programme I & II, 1993–2002 • Economic Recovery Strategy / Results for Kenya, 2003–2007 • Public Service Transformation Strategy, 2010–2014 • Public Service Transformation Framework, 2017–	• Economic Recovery Strategy / Results for Kenya, 2003–2007 • Public Service Transformation Strategy, 2010–2014
Nigeria	• Replacement of APER with AUTOPAS, 2009– • Federal Civil Service Strategy and Implementation Plan, 2017–	• Minister Performance Contracts, 2011–2012
Senegal	• Modernization of the State, 1990–1999 • Programme Nationale de Bonne Gouvernance I & II, 2002–2011 • Reform Master Plans (Schémas Directeux) / Plan Sénégal Emergent, 2011–	• Performance Contracts, 2010–
South Africa	• Legal and administrative reforms (1994–98)	• Senior Management Service and Performance Management and Development System (2001–) • Ministers Performance Agreements (2009–10) • Head of Department Performance Management and Development System Relaunch (2017–)
Zambia	• Public Sector Reform Programme, 1993–1999 • Public service Capacity Building Project, 2000–2005 • Public Service Management Programme, 2006–2011 • Public Service Transformation Programme I & II, 2013–	• Public Service Capacity Building Project, 2000–2005 • Public Service Transformation Programme I & II, 2013–

Source: Author's synthesis.

in this area because it would reduce labour unrest."[12] In 1995, a letter from the Office of the Head of Civil Service to heads of all ministries and departments chastised them:

> By the Circular Ref No. PNDC/SCR/A. 08/15 issued in September 1991, all MDAs were requested to institute a Merit Pay Scheme. Under this scheme, staff of Organisations found to be achievers in their job were to be identified and given awards each year. Thus you were required by the Circular (copy enclosed) to create the necessary conditions for making the scheme effective.
>
> 2. A survey of the MDAs has revealed that the scheme is yet to see the light of day in all institutions in the Civil Service.
> 3. One of the guiding principles of the on-going Civil Service Performance Improvement Programme is the emphasis on performance measurement (output orientation). It is thus reasonable to recognize achievements of officers. It is therefore opportune to get the Scheme under way so that achievers can be rewarded appropriately.[13]

This link did not occur under CSPIP either, and after four further efforts to effectively operationalize or modify this system between 2006 and 2012, the key policy document of Ghana's most recent reform wave (the National Public Sector Reform Strategy) once again listed under its activities: "Introduce a performance-related pay based on a well-designed performance contracting system" and "Develop and institutionalize a non-monetary incentive policy and scheme to motivate and retain high performing public sector workers."[14] In Kenya, Senegal, and Zambia, such policies were included in reform packages covering almost every year from the early 1990s through 2019. Yet, despite these repeated efforts—indeed, as evidenced in part by the need for repeated reform efforts—differentiated rewards and consequences were not consistently linked to individuals' performance in any sustained or systematic way in any of these cases.

These performance management systems for rank-and-file civil servants were paralleled in each country at the level of senior leadership by performance contracts or performance agreements for heads of organizations, other senior managers, and even (in Nigeria and South Africa) ministers. The details of each scheme and the timing with which they were adopted in each country were different—as early as 1997 in Ghana and as late as 2011 in Nigeria—but they shared the same combination of participatorily set performance targets linked to organizational work plans, formal scoring of achievement against these targets, and linking of rewards and/or punishments to these assessments. As with annual appraisal systems, these were often inspired by international experience and sometimes (though far from always) linked to donor projects, and in some countries (e.g., Kenya), they had been piloted with state-owned enterprises prior to their rollout in civil service ministries.

Each of these reform efforts had its own idiosyncratic implementation story (detailed in the appendix), but as with the annual appraisal reforms, they shared a common fate—the failure to sustainably link differentiated rewards and punishments to measured performance. Some of the schemes resulted in perfunctory assessments in which everyone scored well; others managed to give differentiated assessments but were not able to link them to meaningful carrots and sticks, and others collapsed after just a year or two.

I was only able to find evidence of two instances in which significant pecuniary incentives were actually delivered for multiple years. Under Kenya's Performance Contract system for permanent secretaries in the mid-2000s, staff members of the highest-scoring ministry were given a "13th-month" salary bonus for several years.[15] Even this suffered from delays, distortions, and implementation problems, however, and eventually fell from prominence in central government. It was also a group incentive rather than an incentive for individual permanent secretaries, and there were never any explicit consequences for poorly performing individuals or ministries (despite the system's intention for such punishments to exist). In South Africa, the Performance Management and Development System for senior managers reportedly delivered differentiated assessments linked to bonuses in its early years but quickly deteriorated into a situation where "people were just getting our performance increases irrespective of their performance . . . so I didn't think that, you know, overall the performance management system worked very well because there are no consequences for poor performance."[16]

The only other partially successful use of performance contracts was in Ghana during the 2010s. Under the performance contract system that had operated for a few years in the late 1990s, one chief director's contract was reportedly not renewed due to poor performance, but soon after, the entire system was scrapped, and contracts ceased being upheld.[17] A later effort to reintroduce performance agreements for chief directors began implementation in 2013 and was still being conducted annually and delivering differentiated performance scores as of 2019, making it relatively long-lived. But the only incentives attached to these assessments were soft ones like recognition or mainly symbolic rewards despite the intention to link these to explicit rewards and consequences.

Why did these efforts to introduce individual-level performance incentives into civil services keep failing to actually deliver incentives? The answer has to do with their focus on creating an objective and highly formalized system to force change in individual behaviors that are, in large part, informal and unformalizable. To see why, let's first lay out how these systems were supposed to function and then examine how, when, and why they deviated from these intentions.

HOW THEY WERE INTENDED TO WORK

Each of these systems was designed to begin with a target-setting phase at the start of each year. Each individual would meet with their supervisor and agree on a set of targets—tasks, activities, deliverables, outcomes—that they would be responsible for delivering over the course of the year and that were linked to broader goals or deliverables for their team, their organization, and/or the civil service as a whole. In the jargon of management, these targets should be S.M.A.R.T.: specific, measurable, achievable (or attainable, depending on who you ask), relevant (or realistic), and time-bound. Different reforms specified the target-setting process in slightly different ways, but the basic idea was always the same: Establish clear and objectively measurable targets that track the work individuals will be doing during the year and against which their performance can later be measured. As one rank-and-file civil servant in Zambia explained, "That is why we set goals and targets, we need to show if we met the target. We need to prove we shine. . . . It is about time people realized they are being paid for something."[18]

At the end of the year, each worker was to be assessed on their actual performance against their targets, sometimes with interim check-ins or feedback points during the year. The idea was that these assessments would be differentiated. Good performers score well, and bad performers score badly. Since workers' targets were mutually agreed upon with their supervisors and were S.M.A.R.T., these performance assessments were, therefore, intended to represent an unbiased indicator of performance accepted by all parties.

Finally, these performance assessments were meant to be used as the basis for delivering some form of reward and/or punishment to the individual, according to their measured performance. The range of rewards envisioned by these thirty-four reforms included financial incentives like bonuses or piece-rate payments; career benefits like accelerated promotion or contract extensions (for senior leaders who are often appointed on nonpermanent contracts); social

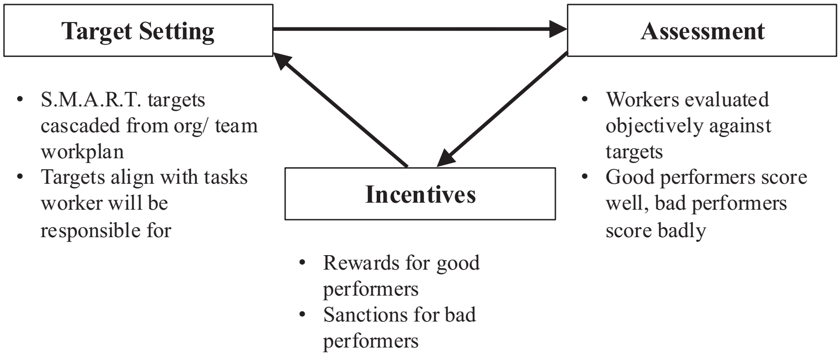

FIGURE 4.1 Intended Structure of Performance-Linked Incentive Systems

Source: Author's synthesis.

recognition like best worker awards or published "league tables"; or sometimes other nonfinancial rewards. For example, a 1991 government circular issued under Ghana's Civil Service Reform Programme suggested "tangible objects, eg. Clock, cloth, wrist-watches, furniture, set of books, radios, scholarship for a child for one year."[19] The range of punishments envisioned was just as broad, up to and including censure, demotion, pay reduction, or dismissal. Though the details and types of rewards and sanctions thus differed, the common thread was that extrinsic incentives to elicit greater effort from bureaucrats through the promise (or threat) of carrots and sticks were the key mechanism through which these annual performance management systems were envisioned to change bureaucratic behavior. Figure 4.1 illustrates this idealized annual performance management cycle.

HOW THEY ACTUALLY WORKED

In reality, however, there were common patterns in the ways that these individual-level performance-linked incentive systems fell short of the aspirations at each stage.

Target-Setting

At the target-setting stage, there was an inconsistency between the ideal of establishing S.M.A.R.T. targets linked to organizational work plans and the reality of how difficult it was to ex ante specify exactly what each individual civil servant should do during the year. The senior public servant and researcher Sylvester Obong'o explained of one performance appraisal reform effort in Kenya:

> What happened with that new system is that not everything deliverable ended up in the appraisal, so the targets were actually set, and some people ended up setting targets on very easy things [to achieve], which are then measured, but you also end up doing a lot of other things which are not actually in your performance contract. . . . Ninety percent of what I do and what I'm engaged in is not part of those targets by [the] nature of the public service . . . it doesn't really make a lot of sense to have these targets at the beginning of the year, which you put two or three, but what you end up doing is not what . . . you plan to do.[20]

An evaluation of Zambia's performance appraisal reform efforts also highlighted the inherent unpredictability of much of the work that civil servants do, particularly in the types of policy and oversight roles prevalent in the core civil service: "The best laid work plans can be de-railed when urgent and pressing work duties displace work plan targets. . . . Political directives from above, and outside of the scope of the work plan, must be recognized as part of the working culture."[21]

A further issue is the disjuncture between individual effort and team or organizational performance that arises in contexts where team production is prevalent, like most civil service settings. The scholar Danny Sing describes how this affected South Africa's Performance Management and Development System (PMDS) appraisal system for managers: "Another concern that emerged was the appraisal process may not reflect adequate correlation between individual performance and overall organizational performance. It is generally accepted that an outstanding rated performance of an HOD [heads of departments], means, that he/she is leading an organization which performs optimally. However, the PMDS does not provide an instrument to deal with the potential disjuncture between individual performance and organizational performance."[22]

The intended linkage between individual targets and organizational work plans also created problems when the work plans themselves were flawed, incomplete, or even absent. For example, when performance contracts were first introduced in Zambia under PSCAP in the early 2000s, permanent secretaries lacked not only annual work plan targets but also basic job descriptions.[23] A subsequent effort to reintroduce performance contracts in 2015/2016 fixed this by creating clearly delineated schedules of targets each year that were largely extracted from the ministry's work plan, which in turn came from the National Development

Plan.[24] However, since the National Development Plan itself was too ambitious, each permanent secretary's target was unrealistic, and so almost all of them scored poorly on their assessments.[25]

Similarly, for rank-and-file staff, it was often difficult to create individual targets that were both linked to organizational work plans *and* could be used for individual performance measurement since many important and measurable actions or outputs depend on team production and/or complementary inputs from other teams. But, on the one hand, targeting these directly risked the individual being measured as low-performing due to the inaction of others (or vice versa). On the other hand, focusing on more narrow and individualized targets risked ignoring the individual's contributions to team efforts, which in the civil service can seldom be reduced to the sum total of prespecifiable individual tasks.

These technical challenges, combined with workers' understandable desire to avoid poor performance assessments—particularly when these ratings were intended to be linked to rewards or punishments—pushed many workers toward setting targets that were vague, easy, or soft. One South African civil servant explained of senior managers' appraisal targets:

> The way they design it is that it's not something that comes back to them . . . for example, a simple one would be you need to build X number of houses per year, so you receive a budget of X billion rand, you need to build so many houses, the manager was simply right there to oversee the building of houses so whether we build ten when we were supposed to build twenty, I have overseen the building of the houses. I didn't put a target on building 20 houses although I received funding for 20 houses and therefore when you do the assessment, [you can say] " . . . I did oversee it, these are the reports. . . ." So it's the manipulation of the system to a large extent.[26]

These dynamics manifested themselves even in relatively more successful cases, such as Kenya's widely hailed use of performance contracts in the mid-2000s. Some officials perceived that the incentives built into the system pushed organizations over time toward setting easy targets. Others reported that there were "a lot of accusations about soft targets" in centralized ministries with administrative remits, whereas service-delivery-oriented ministries, such as health or agriculture, faced targets that were more tangible and harder to affect.[27] Researcher Abraham Muriu reports that a government-appointed expert review panel in 2010 found that the "setting of targets had not been well coordinated and that the [performance contracting] process was not in tandem with the budget process hence impeding on performance improvement efforts."[28]

This litany of failings in the target-setting process boiled down to one root problem: Important parts of what civil servants do cannot be fully and objectively specified in advance. This might be because some tasks require contextual judgment, because they are hard to anticipate, or because they must respond to changing circumstances or actions of other colleagues and stakeholders. There are, of course, better and worse ways to handle these technical challenges and more and less serious ways to approach the target-setting process. But as long as at least some important aspects of performance are not fully verifiable, the ex-ante setting of formal targets is necessarily incomplete. And as we examine next, this incompleteness of targets—indeed, even the perception of it—undermines the ability to use them to assess performance.

Assessment

The dominant empirical pattern that unfolded at the assessment stage of individual-level performance incentive systems was a lack of differentiation of performance ratings. Most commonly, almost everyone scored highly; less commonly, nearly everyone scored poorly. Either way, there was little differentiation in measured performance. How and why did this happen?

First, the fact that even the most precise targets only specified a fraction of what each individual was responsible for during the year meant that individuals were de facto expected to undertake many actions that were not prespecified but were important for their own performance, their team, and their organization. But since these tasks were not captured in their formal targets, their performance on them could not be rated in the same way as the prespecified tasks. Nor was it possible even to define what percentage of an individual's work comprised prespecified versus unforeseen tasks. Sticking rigidly to considering only prespecified tasks when giving performance ratings would risk undervaluing these unformalizable, nonverifiable tasks and distorting individuals' effort away from them—at the expense of overall performance. As one South African public servant remarked, "In the public service, where what you're actually trying to achieve is much more nebulous and harder to define, it doesn't work very well."[29] Most supervisors (or central rating authorities, in the case of senior leadership performance contracts), therefore, erred on the side of generosity in their assessments.

Second, supervisors seemed to recognize that individuals' performance against their targets often depended on the provision of adequate inputs, on the completion of complementary actions by other people, or on other factors outside the individual's control. The most obvious manifestation of this is related to the provision of the financial resources needed to undertake many activities. Organizational work plans were often underfunded, and even when budgetary provision was made, the promised funds were often not actually released during the year. For example, one expert involved in Zambia's PSRP reform explained that individual targets were usually taken from organizational work plans but since the Ministry of Finance frequently gave ministries budget ceilings of only 65 percent of the cost of these work plans, it was inevitable that many activities would never be completed—how, then, could an individual be blamed for not meeting their targets?[30] In Nigeria, the development practitioner and former senior civil servant Joe Abah explained that a performance contract system for ministers was discontinued one year after it was found that "not one minister met the targets that they had agreed to" and described how ministers objected to the system's premise. "How can we meet these targets when you didn't release all the money for the budget, and we have no control over our staff . . . we can't hire and we can't fire, so how can you hold us accountable to something that we have no control over?"[31] In several contexts, attempts were made to address these issues by creating a section of the evaluation that listed mitigating factors or that released individuals from their obligations if the government did not provide adequate resources, but given the inherent uncertainties of government fiscal management, this tended to further undermine the perceived objectivity of the ratings and thus the ability to give differentiated performance rankings.

Third, while supervisors might have had a good sense of how well each officer was performing against their responsibilities, proving it in an objectively verifiable way was challenging except in cases of serious malfeasance or law-breaking. This wasn't a problem of information per se. Among my interviewees at various levels of seniority, it was widely agreed that everyone within a team basically knew who was a good or bad worker. This makes sense in the context of core civil services. These are people who interact and work together every day, mostly in the same office buildings. They know who has the best technical expertise, who always turns memos around promptly, and who can't be relied upon for important tasks. But compiling the evidence trail needed to justify a poor performance rating or a sanction to a third party was difficult and time-consuming, and in cases where important dimensions of performance couldn't be or weren't prespecified, it could be impossible. Supervisors thus often shied away from giving low scores even to individuals they knew were underperforming. As one senior public servant in Ghana remarked on performance appraisal scores, "Everybody is very very good, but you and I know that when it comes down to productivity, not everybody is excellent."[32]

Efforts to improve the rigor of performance reporting had the perverse effect of reinforcing the incentive (discussed above) for individuals to set targets that were less meaningful but were under their sole control: "We went through a phase where people were trying [to focus targets on results and outcomes] but the Auditor General also started auditing our performance data and expressing concerns about whether our performance data was also reliable and accurate and all that. And that made everyone go back to input and process targets. They were very strong on the SMART principle, they were using that in doing the audits."[33]

Finally, the nondifferentiation of performance ratings was also due, in part, to a misalignment between the systemic benefits and private costs of having differentiated ratings. The costs of giving bad performance ratings fall entirely on the supervisor doing the rating, while many of the benefits of having functioning and differentiated performance ratings are diffuse and system-level. Put yourself in the shoes of a supervisor who is considering giving a subordinate a poor performance rating. It will probably be an unpleasant conversation. They are likely to perceive your rating as unfair, subjective, and potentially biased. They may react not by working harder next year but by becoming demotivated and creating negative dynamics within the team. They might even complain to their union, the media, an opposition party, or even your own bosses that you are persecuting them due to their political allegiance, ethnicity, or as a result of your own wrongdoing—an allegation that would be harmful to you regardless of whether or not it was true.[34] In contrast, the biggest benefit (from a system perspective) of giving them a poor rating would be that the central authorities know not to promote them and other managers know not to offer them transfers into their teams, so you are likely to be stuck with that individual even longer. Supervisors, thus, have little strategic incentive to give bad performance ratings. As one former South African civil servant explained, supervisors "shy away from any form of conflict and just do the tick-box exercise . . . [they think] 'it's not my problem, it's somebody else's problem' and they managed it on that basis and it just goes away, nobody bothered."[35]

Incentives

It's easy to see why having nondifferentiated performance assessments makes it impossible to give differentiated performance incentives. But it's also worth considering why it is hard for bureaucracies to actually give out rewards and punishments even if individuals do have differentiated performance ratings.

Individuals have various options to resist if they feel that they are being punished unfairly or that they also deserve rewards that others are receiving—which is to say always, given the difficulties of perfectly prespecifying targets and then objectively proving performance. As discussed above, they can appeal to public service commissions; file suit in courts; complain to unions, the media, or opposition parties; and take other actions that create costs both for individual managers and for the government as a whole. And the stronger the incentive, the stronger the resistance it will provoke: No one is likely to complain about a nonfinancial recognition award to a "best worker," but if significant money or the continuation of their job is on the line, many workers will claim—rightly or wrongly—that they are being unfairly and subjectively persecuted and take whatever steps they can to resist.

Furthermore, it is common in civil services in Africa and worldwide for individual civil servants to have connections to other powerful figures, both in higher ranks of the bureaucracy and in political offices. These connections represent an additional avenue through which individuals can contest the allocation of incentives. For example, in Ghana, there was reportedly one instance in which the Office of the Head of Civil Service tried to terminate the contract of a chief director who had been assessed as performing poorly, but that individual made a direct appeal to the Office of the President and was able to secure a contract renewal.[36] Similarly, one South African former civil servant explained: "But there's also the political influence, the moment you start taking action against individuals, there's also—because there's links to politicians [of those] who got appointed through the politicians. . . . So those played a role as well, so therefore the moment you initiate a process . . . you end up with political interference to some extent. Or even if it's not political, you still get administrative interference from higher up the chain. So managers then sit back and say 'but why do I need all those stress[es] in life' so you just rather not get involved."[37]

A related challenge for following through on performance-linked incentive schemes arose when the institution or individual that appointed them had different priorities from those that were written in organizational work plans or official performance targets. For example, a minister might have different personal or political objectives than those laid out in a ministry's medium-term plan, especially as ministers come and go or political situations change. This challenge was especially acute for senior civil service leaders, who sit in the middle of the political-administrative interface—not only responsible for carrying out work plans and administrative processes but also expected to be responsive to the political priorities of their ministers. While these two roles are intended to coincide, they can often diverge. When a manager scores poorly on their formal performance metrics but the political leader who appoints them is happy with them (or vice versa), then there is a natural tendency for the preferences of the political leader to win out. Managers know this, which undermines the credibility of the performance contract. As one retired Zambian civil servant reflected, if permanent secretaries' main loyalty is to the authority that appointed them, then how can a performance contract be anything more than symbolic?[38]

Interestingly, interviewees across all contexts reported that it was not impossible or even uncommon for individuals to be fired or disciplined. However, this was generally only for cases of severe or criminal misconduct rather than poor performance. Individual failings in such cases were highly verifiable: Regulations and codes of conduct provide relatively clear and complete ex ante specifications on what to do or not do, and many forms of criminal or financial malfeasance are ex post provable to third parties. Of course, not all instances of misconduct were caught or punished. But the possibility of levying strong sanctions for such infractions in at least some cases stands in contrast with the near-universal inability to do so for reasons of poor performance. As one former South African civil servant commented, "There are no consequences for poor performance . . . it was always difficult to dismiss people on the basis of poor performance. You could do so on misconduct but poor performance it was very difficult, so you just keep them in the system."[39]

In addition to provoking resistance, efforts to attach strong incentives to performance also tended to distort earlier stages of the performance management process by encouraging individuals to set easily achievable targets and increasing the pressure on managers to give positive assessments. For example, with Kenya's performance contracting system in the mid-2000s, once rewards began to be introduced, "people started to look for easy targets where they could score highly and then be rewarded," often by setting targets related to carrying out processes rather than to the ultimate impact of their actions.[40]

Because of all these challenges in delivering differentiated incentives, most such schemes either dropped the incentives completely or delivered them in a largely nondifferentiated fashion. For instance, a consultancy report on Zambia's efforts to operationalize its APAS annual appraisal system during the PSRP and PSCAP reforms described this outcome in devastating fashion: "As time has passed the real purpose of the APAS report has become the justification of pay increments and promotions. This has led to the a [*sic*] view that completion of the form is a necessary evil to which one should devote as little time and thought as possible. The result in many instances is a report replete with inconsistencies, contradictions and very little assessment of performance that bears little relation to a real work plan and virtually none to the organisational and strategic plan."[41]

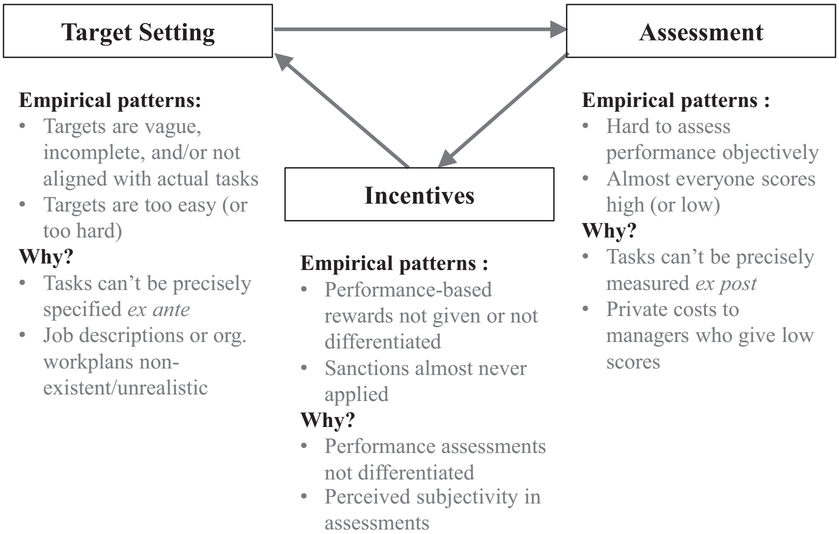

FIGURE 4.2 Actual Operation of Performance-Linked Incentive Systems

Source: Author's synthesis.

Similarly, in South Africa, one public servant remarked, "We got into a phase where you were almost guaranteed a performance bonus. . . . And there wasn't much of a correlation between unit or department performance and performance bonuses. Performance bonuses became a 13th check almost, it was expected."[42] Figure 4.2 summarizes the most common ways in which the actual implementation of these performance management systems diverged from how they were intended to operate.

PERFORMANCE-LINKED INCENTIVES OFTEN FAIL IN HIGH-INCOME COUNTRIES, TOO

In my interviews with civil servants and donors alike, I often encountered a pair of common perceptions about performance-linked incentive reforms. First, that such systems frequently existed and functioned properly in high-performing civil services in rich countries. Second, that their implementation failures in African countries were due to contextual factors like nonindividualistic or nonperformance-oriented cultures, political interference, or low bureaucratic effectiveness. In other words, it was often perceived that these systems were the kinds of things that high-performing civil services did, that doing them was the way to become high-performing, and that failures and challenges were due to historical, social, or political factors specific to Africa.

While a global review of performance-linked incentives is far beyond the scope of this already broad book, even a cursory look at how such systems have tended to operate in high-income contexts challenges these perceptions. (As an aside, the widespread assumption that richer countries always have more effective government bureaucracies than poorer countries is also contentious, but that's an argument for another book.) It also helps illustrate this chapter's main argument—that the root of implementation failure for these systems is their focus on creating formal, objective, and mechanistic links between measured performance and rewards or sanctions.

A good place to start is New Zealand, whose experiments in the late 1980s and early 1990s with public sector performance contracts (along with a host of other New Public Management–style reforms) were often seen as a model and provided much of the initial inspiration for their global spread. For instance, as early as the late 1990s, the scholar Allen Schick famously discussed the global influence of New Zealand's reforms on reform thinking in developing countries.[43] The archived draft minutes of a 1993 meeting of the steering committee of Ghana's Civil Service Reform Programme (CSRP) provide direct evidence of one attempt to import ideas and experience from New Zealand into reform planning in Ghana, with the minute-taker recording that one foreign consultant representing the World Bank at the meeting "said that he could see some similarities between the circumstances facing the New Zealand Civil Service and the Ghana Civil Service" and another stating that "it would be advisable to look at what had been done in the U.K. and New Zealand."[44] Similarly, South Africa's 1997 Presidential Review Commission report discussed having referred to documentation on New Zealand's reform experience that was "particularly relevant to its work."[45] And Zambia's performance contracting system for permanent secretaries in the early 2000s was reportedly based, in part, on a suggestion by an expatriate consultant from New Zealand.[46]

But the actual details of New Zealand's performance contracting system differed in important ways from the image of it that traveled around the world and the performance contract systems that were based on it, at least in these six countries. These details are contained in a number of contemporaneous documents and academic studies, but to better understand the original intent of these systems, I also interviewed Ian Ball, who worked as a civil servant in the Treasury during this period, helping to design and implement these reforms and later worked as a consultant and academic.

First, Ball emphasized that the term *contract* was a "metaphor"—the contracts were never intended to be enforceable. Indeed, as Ball emphasized (and documents confirm), they were almost always called performance "agreements" rather than contracts, although the latter term is sometimes used as a verb (*contracting*). Although many governments around the world took this metaphor literally and tried to create systems that were objective and enforceable, Ball explained why such literal interpretations were misguided:

> We knew we operated in a complex world and nothing was as simple as a fully specified contract anywhere, whether the private sector or the public sector, so we didn't expect to be doing that. . . . The idea that you would have a formulaic relationship between services and remuneration and that alone would be what you meant by chief executive performance is kind of, to me, is very simplistic, is not what we were trying to do. We were trying to create a management system where the whole system—formal and informal—would work with the personnel and financial management elements working in harmony.[47]

Second, while performance contracts in many of the cases I studied were criticized for focusing on intermediate actions rather than final outcomes like service delivery, in New Zealand, "there was a deliberated decision not to seek to use outcomes in order to define the accountability of chief executives and their departments. The reason for this is that the individual chief executive very rarely has sufficient control over outcomes to make accountability effective."[48] Ball explained:

> The other thing that was very much in our minds in thinking about how this would work was that any system works with formal components but also informal components. So for example, in relation to outcomes, you would want the system to be working in such a way that the Chief Executive was concerned not just with the delivery of the pre-specified outputs, but also with whether or not those outputs were adding value in terms of the outcomes that they and the ministers were trying to achieve. . . . We explicitly were regarding outcomes as the ultimate rationale for public actions, but saying you can't contract those or reach agreements for those in the way that you can for outputs, so you would have wanted the Chief Executives also to be contemplating whether the particular set of services was contributing to the outcomes that the agency was seeking to achieve—and other informal components of their performance.[49]

Enforcement challenges also motivated this focus on outputs rather than outcomes, with Ball noting in a 1992 conference presentation that "accountability for outcomes was rarely made to stick."[50] As a 1996 review of reforms by the scholar Allen Schick commissioned by the government of New Zealand emphasized, "The focus on outputs is in contrast to both its previous input-based appropriations and the outcome-oriented system favoured in much of the management reform literature."[51]

Third, the main envisioned driver of improvement from the performance agreements was the discussion and clarity that they would give rise to, rather than the carrots and sticks. As Ball wrote in a 1994 journal article: "The emphasis placed on contracting before a period begins, as distinct from measuring performance after it ends, is another key feature of the reforms. Our emphasis has been heavily on the former . . . a focus on ex ante contracting, rather than ex post performance measurement, has been hugely helpful in clarifying the roles and performance expectations of chief executives, and has provided much better focus to departmental activity."[52]

At the same time, even this emphasis on ex ante specification was observed to sometimes cause managers to approach it as a perfunctory exercise: "The focus on ex ante specification has sometimes led to a checklist mentality which is positive from the perspective of having managers take accountability seriously for tasks they are expected to complete. It is less desirable if it narrows responsibility to simple compliance with what is on the list, and prompts Chief Executives to disregard responsibility for items not specified."[53]

Fourth, the evaluation of performance and delivery of associated incentives was not objective, transparent, or highly differentiated. Rather—and by design—it contained important elements of subjectivity and judgment, was largely confidential, and seems to have led to relatively little actual differentiation in extrinsic rewards. Ball recalls that the actual evaluation of performance by the State Services Commission "was treated as a personnel, in-confidence kind of issue" rather than being done in public and that he and his colleagues "would never have expected you could put that into a formula, that there would be a formula which would say 'this is the relationship between the chief executive's income and the service delivery.' "[54] And even with this subjectivity and nontransparency, there was little differentiation of rewards or punishments for performance—although judging this involves some speculation, as information on it was never made public.

> The understanding that I had at the time is that . . . 90 percent of people got the same [rating and remuneration]. That the SSC was very reluctant or unable

> or unwilling, whichever it was, to differentiate in a way. I suspect it was that they weren't confident in differentiating in a way that they could support and justify. . . . I know within the Treasury we felt that the State Services Commission was very reluctant to determine that one Chief Executive would get a significant performance component in their remuneration and another would arguably even get a reduction in their remuneration.[55]

Similarly, Schick's 1996 review noted the widespread perception that the State Services Commission "has been reluctant to dismiss weak performers or to use pay differentials to reward strong managers and penalise weak ones. . . . It is unrealistic to expect SSC to approve large pay differentials in response to differences in chief executive performance. . . . In practice, pay differentials tend to correlate more with departmental size than with performance."[56]

Thus, New Zealand's system of performance agreements was far less focused on formalizing every aspect of the process and on leveraging carrots and sticks than most of the systems it inspired around the world, including in the six countries covered by this book. This is not to imply that New Zealand's system in the late 1980s or 1990s was perfect or even necessarily effective—again, this is a debate for another book[57]—or that countries should try to imitate it. Rather, my point is that even one of the key global archetypes of individual performance-linked incentive systems in government was far less focused on forcing change through formal incentives than is typically imagined. It was designed this way because of many of the same challenges of target-setting, performance assessment, and incentive delivery that undermined the performance contracting and performance appraisal reforms in the six countries described in this chapter.

Many of the same patterns are present in other high-income countries' governments. For example, in the United States, a 2016 report found that 99.7 percent of US federal civil servants were rated excellent *or above* on their annual staff appraisals.[58] While many interviewees in the six African countries studied in this book attributed the nondifferentiation of performance ratings at least partially to cultural factors, the fact that there is so little differentiation in a context like the United States—often considered to have a highly individualistic culture in which economic incentives are prevalent and widely accepted—ought to cause us to be skeptical that national cultures are the main constraint on such systems' operation. Similarly, a study of local governments in Italy found that performance ratings exhibited very low differentiation both because objectives are hard to define ex ante and differentiated ratings ex post "may ruin the internal climate," with the result that "additional money is often not seen as a recognition for achieving a predefined set of objectives, or more generally for a superior performance, but it is taken for granted as additional salary."[59] Indeed, a 2005 report by the OECD found that across its member states: "there is often a gap between the stated existence of a so-called 'performance-related pay scheme' and its concrete functioning, which may be barely linked to performance. . . . Performance pay is an appealing idea, but the experiences reviewed in this study indicate that its implementation is complex and difficult. . . . Performance measurement in the public sector requires a large element of managerial judgement. The notion of performance itself is complex, owing to the difficulty of finding suitable quantitative indicators and because performance objectives often change with government policy."[60]

It further noted that "highly detailed and inflexible performance criteria and rating frameworks, though they can be reassuring to managers and managed alike, are often detrimental to the smooth running of a PRP [performance-related pay] scheme" and reported the findings of an earlier OECD report that "in most PRP schemes in use in the 1980s, more than 95 percent of managers were rated as 'fully satisfactory or better.' "[61] These findings echo much of the literature on performance management at the organizational (rather than individual) level in OECD countries by scholars such as Beryl Radin and Donald Moynihan, who argue that overly formal and mechanical approaches to managing agencies' performance fit poorly with the complex reality of large bureaucracies in democratic contexts.[62] Of course, this is not to say that performance-linked incentive systems always fail, nor that good performance always goes unrewarded. Indeed, evidence from the Global Survey of Public Servants reveals significant variation both across and within countries in the percentages of public servants who say that their performance matters for pay rises and/or career prospects[63]—although the survey does not explore the extent to which that linkage is due to the type of highly formalized performance-linked incentive systems examined in this chapter as opposed to more flexible and discretionary mechanisms.

Taken together, then, the frequent difficulties encountered in successfully operationalizing performance-linked incentives in civil services around the world pose a deep challenge to explanations for the nonimplementation of performance-linked incentive reforms that rely on contextual factors unique to African countries or to low- and middle-income countries. Rather, many of the same patterns appear even in high-income countries with very different contexts, resources, and constraints. Instead, this evidence reinforces the interpretation that the nonimplementation of these reforms stems from their overfocus on trying to force behavioral changes through formal systems and incentives.

OTHER TYPES OF REFORMS ALSO FOCUSED MAINLY ON CHANGING FORMAL RULES AND STRUCTURES

The formal focus of reforms was not restricted to performance-linked incentive programs alone. Rather, it was a consistent feature of almost all the reform efforts undertaken in these six countries over the three-decade span I studied.

This is most obviously the case for the category of reforms I term "salaries and structures," which focused on organizational restructuring, staff redundancies, and changes to staff pay scales. Even more than the performance-linked incentive systems, these reforms focused exclusively on changing formal structures and employment contracts with the idea that these would lead to better performance. Indeed, they were often paired with—or viewed as predecessors to—performance-linked incentive reforms. The idea was typically that downsizing would create fiscal space to enable better pay for high performers or that decompressed pay scales would give staff more incentive to perform well in order to be promoted. But the linkage of these changes to measured performance never arrived. For example, Ghana's Single Spine Pay Policy reforms of 2006–2011 envisioned first harmonizing and improving pay to provide the basis for a subsequent reform that would explicitly link performance to pay, but by the time the new pay scales were adopted, the cost had ballooned out of control, years had passed, and there was no appetite for the envisioned incentive reform. Across countries, despite the myriad shortcomings of existing pay scales and organizational structures, formal salary and structure reforms typically yielded no obvious performance improvements and ended up saving far less or costing far more than was anticipated.

Organizational management and capacity reforms sometimes recognized the importance of informal practices for performance, but then tended to revert to a focus on formal structures and processes in how they actually tried to make change. Performance improvement funds in Ghana and Zambia accurately identified the potential for bottom-up improvement based on ideas generated in a decentralized fashion, but the support they then offered to these ideas was in the most formal of currencies within government bureaucracies: budget allocations, with all the spending and reporting requirements they entail. Similarly, most organizational performance review mechanisms focused overwhelmingly on objectively verifiable indicators of performance, whether in terms of activities, outputs, or outcomes. This focus on formalizable processes and measures in organizational performance management systems encountered the same implementation challenges, behavior distortions, and limitations as did their individual-level counterparts described above.

South Africa's Management Performance Assessment Tool (MPAT, which operated from 2011–2016), under which detailed audits of the implementation of a range of management processes were undertaken each year, is, in some ways, the exception that proves the rule. Attuned to the risk of overfocusing on the formal, the designers of the MPAT assessment created a performance rating scale of 1–4 for each process, in which scores of 1–3 represented various degrees of noncompliance or perfunctory compliance with the formalizable aspects of required and recommended organizational management processes. The highest score of 4 was reserved for organizations that demonstrated that the process was not only being followed in the letter of the law but was also being used to improve performance—in other words, for consummate performance.

This effort to encourage such unformalizable behaviors was a promising and unusual part of the MPAT reform. However, it was undermined to an extent by the system's intention to use the resulting scores as accountability tools and metrics for delivering incentives. This required a burdensome process of seeking documentary evidence for all these behaviors—which itself ensures that many unformalizable behaviors will be missed—and also led to distortions. As one civil servant close to the scheme explained,

> People started learning the system and started playing us, because then it became about the rating and the scoring. We did initially introduce competition and give awards out and say who's the best and who's the worst to try to motivate people to make the change, but unfortunately it then became about the score. So we tried to give awards about who's the most improved department, but it became a lot about the scores, not about "am I improving," "are we getting better." . . . Then we started getting a huge amount of pushback from the departments and the [directors-general] and that, saying "why are we focused so much on compliance and we should rather focus on outcomes and all that."[64]

As with the individual-level performance-linked incentive reforms, the focus on assigning formal scores to use as the basis for doling out carrots and sticks undermined the learning potential of the participatory diagnostic process, discouraging line ministries and central management agencies alike from engaging openly and frankly in discussions of performance and how to improve it. The case of the MPAT thus illustrates that even organization-level performance rating systems could trigger backlash when used for accountability purposes: "It didn't go well with the ministers and the departments that were at the bottom. We also went public with the results which put further pressure. And politically it wasn't liked by some, the approach, and that's kind of why it died a quiet death, because maybe we were too transparent and pushed too much."[65] In the same way as with individual-level performance incentives, organization-level performance management systems collapsed when they tried too hard to force changes in behavior through formal incentives and accountability systems.

Within the category of service-delivery-focused reforms, the most common intervention was the creation of new customer-facing units charged with making it easier to access services, such as "one-stop shops" or client service units. In most cases, these were also formal-focused ways to try to solve complex organizational problems by creating a new unit with a new name, new physical infrastructure, separate budget line, and separate staff. And in most cases, these new units showed initial promise but fell into disuse once the donor project that funded their construction ended or the political leaders that launched them moved on. Kenya's Huduma Centers, launched as one-stop decentralized service access points in 2013, are perhaps the most salient exception: Widely viewed as fairly successful (even if still facing some challenges), the main difference between them and other similar reform efforts lay not in their formal design but in the sustained and gradual improvements in services offered, reliability, and performance monitoring and feedback in the centers. Thus, the creation of formal systems, structures, and processes in itself is not bad; indeed, it is often a necessary part of reforms in large bureaucracies. But the creation of these formal systems alone isn't sufficient for them to change people's behavior in the ways they generally aim to.

Once again, this pattern of focusing reform efforts on formal rules, structures, and processes—and the limitations of doing so—appears in reform experience and evidence beyond these six countries. To cite just a few examples: Matt Andrews notes the widespread tendency of donor-supported institutional reforms in developing countries to focus on the form rather than the function of institutions;[66] Kate Bridges and Michael Woolcock observe that 92 percent of indicators of public financial management reform projects in Malawi focus on formal "regulative" processes "targeted at shaping behaviour through the threat of sanction";[67] and Jeffrey Braithwaite comments in his review of health reforms in high-income countries that, "The boxes on the [UK National Health Service] organisation chart have regularly been redrawn to little benefit. Although such reorganisations do produce structural change, they do not greatly alter entrenched cultures, much less downstream clinical outcomes," with similar evidence in Australia and other countries.[68]

In emphasizing the tendency for reforms to focus and rely upon formal changes, I don't want to caricature these reforms or their designers. Many reforms did contain important informal elements, many reformers recognized that formal changes didn't automatically translate into practice, and there were a handful of reforms that consciously focused on achieving unformalizable changes, whether through informal interventions or by leveraging formal processes to support informal changes. I explore these in more detail in chapters 6 and 7 as a way of exploring alternative approaches to reform. But such efforts were usually secondary to the dominant focus-on-the-formal approach, which tended to predominate in the vast majority of reform efforts. Ultimately, the success of a reform depends on the interaction between its formal and informal aspects, which must be deliberately designed to complement and reinforce each other—a goal that requires reformers to think beyond forcing compliance with formal processes as a mechanism for positive change.

• • •

In chapter 3, I argued that reformers' ideas about performance and bureaucratic change—their perceptions, mental models, and understanding of evidence and practice from other contexts—are key determinants of the adoption and design of reforms. The same is true of why so many reforms focus on performance-linked incentives: The conceptual logic is direct and intuitive, and many people think that such systems are widely successful in other contexts.

The evidence in this chapter should lead us to be skeptical not just of claims that performance-linked incentives are necessary to improve performance but also of their viability as a tool for sustained improvement when adopted at scale in civil services. This is certainly true of their track record of implementation in the six countries I study, and similar challenges have been experienced in many high-income countries as well. More broadly, bureaucratic performance is simply too hard to specify and too hard to measure—too nonverifiable—for positive behavior change to be forced by reforms that focus on changing formal rules and processes alone.

This failure to deliver differentiated incentives across these thirty-four reforms ought to be striking not only for reformers but also for researchers. The impacts of performance incentives in government are the subject of an entire cottage industry of studies by academics and policy researchers, making them one of the most intensively researched reforms for improving service delivery around the world. However, in order to isolate the causal impacts of the incentives on performance, these researchers usually study the rollout of incentives in carefully controlled settings: randomized control trials, donor projects, and small-scale pilots in specific sectors. These studies have mixed results in terms of their impacts on service delivery,[69] but this chapter's findings suggest a more fundamental problem for such policies: They seem to be almost impossible to implement and sustain at large scale outside of carefully controlled or limited settings in many (perhaps most) contexts. There have been a handful of economic studies on high-profile implementation failures in scaling pilot performance-linked incentive reforms,[70] but as with qualitative case studies of such reforms, these have often been viewed as examples of isolated and idiosyncratic challenges rather than part of a broader pattern.

While I hope that this chapter has provided new evidence and helpful ways to think about these issues, these insights also reflect the thinking of many thoughtful civil servants and reform designers. To cite just two examples, in his 2014 analysis of the failures of past reform efforts in Nigeria, Tunji Olaopa cited reforms' "Overemphasis on changes in structures and procedures in disregard for the most critical and challenging soft side of culture change that enables shift in business behaviour."[71] Similarly, Joe Abah explained his time as Director-General of Nigeria's Bureau of Public Sector Reforms, "It's actually something we've actually always argued throughout all of the work of performance management while I was at the Bureau . . . performance management, and performance contracting and tying performance to incentives has always failed. In Africa at least it has. People attacked me for it at the time but nobody could show me any evidence of sustainable success from anywhere. So we've always been very aware . . . not to tie [performance] to remuneration, not to tie it to pay, period."[72]

Abah carried on to lament that "this fixation on this use of force rather than meeting of minds continues until this day. And that has been the challenge."[73]

5
The "How" of Reform

Projectization and Its Consequences

The second main mechanism of failure that recurred across reforms was related to the process through which they were designed and implemented rather than the changes they tried to make. In other words, the mechanisms of reform failure weren't only about the "what" of reform but also about the "how."

How these reforms were designed and implemented was, of course, highly varied. As chapter 3 and the appendix discuss, some reform efforts were driven by high-profile political initiatives with great fanfare and others by bureaucratic processes that almost no one outside the civil service was even aware of. Some were initiated and funded by governments, others by donors, and others by both. Sometimes, they were inspired by reforms conducted in other countries; other times, they were homegrown.

But one common feature across the majority of reforms was that they were designed and implemented as one-off projects. They tended to diagnose a performance problem with the civil service system, propose some kind of intervention into that system, and posit that this intervention would "fix" the problem once and for all. Bureaucratic reform was typically viewed as a discrete and often time-bound action: the passing of a new law, the implementation of a three-year plan, or the creation of some new bureaucratic process. This mental model of reform as a one-off intervention—the *projectization of reform*—constituted the second main mechanism of reform failure.

Projectization shaped reforms in several related ways that made it more difficult for them to spur broad-based improvement. In the design phase, it created incentives to exaggerate the potential benefits of reforms, setting them up to be viewed as failures when they inevitably fell short. It also reinforced the tendency for reforms to focus on making changes on paper to formal structures and processes rather than changing actual behaviors—especially nonverifiable ones. In implementation, viewing reform as a one-off intervention undermined the belief among civil servants that the changes being made were here to stay and, thus, undermined their ability both to improve compliance with existing processes and get bureaucrats to undertake important but nonverifiable actions. Projectization also encouraged reforms to be led top-down and be more closely identified with particular political and bureaucratic leaders, potentially increasing not only the pace of change in the short term but also decreasing broader buy-in and sustainability in the medium and long term. Approaching reforms in a projectized way thus undermined their ability to improve bureaucratic behavior and get closer to consummate performance.

This chapter starts by defining *projectization* as I use it, giving examples of various ways it manifested in civil service reforms in the six countries covered by this book and linking it to existing research on organizational change in these and other contexts. It then explains how the mechanism of projectization shaped five aspects of reforms:

1. The *expectations* or goals of reforms
2. The *content* of reforms—i.e., the changes they aimed to make
3. The *implementation* of these measures
4. The *leadership* style of reforms
5. The *politics* around reforms

If projectization was a mechanism of failure, then these were five submechanisms that comprised it in various combinations. Together, they undermined reforms' ability to generate sustained improvements in civil servants' day-to-day behaviors. The chapter closes by recontextualizing these arguments about projectization in the broader picture of reform and discussing whether and when some degree of projectization might be appropriate. It also begins the transition from trying to understand why so many reforms failed to discussing the ways they succeeded—and how they could be even more successful.

WHAT IS PROJECTIZATION?

As an approach to reform, the core feature of projectization is its one-off nature. Projectized reforms perceive a civil service system that is in an undesirable state;

design and deliver some kind of intervention into that system; and expect that once that intervention has been made, the system will operate in a more desirable state. Projectized approaches to reform reflect the influential model of behavioral change and organizational development proposed by psychologist Kurt Lewin in 1951. Lewin argued that planned change unfolds in three steps: the unfreezing of existing undesirable patterns of behavior, making desired changes to these behaviors, and refreezing these new behaviors into new patterns. In this "episodic" model of change, reform is a discrete intervention into an otherwise stable system.[1]

In the realm of civil service reforms, projectization also often takes on several other features, including the following:

- **The separation of reform activities from core organizational processes**. The process of designing and implementing change is seen as something special, unusual, and exceptional and is carried out separately from the routine, repeated, day-to-day tasks of service delivery and administration. Often (but not always), reform activities have their own acronyms, budget lines, and/or implementing teams. Reform is something that is done *to* the organization's core tasks and processes rather than *as part of* them.
- **A clear distinction between the design phase and implementation phase of reform, often resulting in the predefinition of activities and outputs**. Reform is designed by a distinct and dedicated team of senior leadership, consultants, and/or a specific unit. Staff and stakeholders outside this team may be consulted or "sensitized" about the reforms but only for inputs into this otherwise separate design process. The design phase specifies a set of changes, activities, and outputs that are then to be executed or delivered as part of the subsequent implementation process, often with little scope for flexibility or adaptation.
- **An envisioned end to the reform process**. In some cases, projects have predefined start and end dates (e.g., three-year donor projects or four-year reform plans). In other cases, the reform ends once the envisioned change or output—the passing of a new law, the establishment of a new organizational process, the conducting of a set of trainings or organizational performance reviews—has been completed.

Thus, projectization is an umbrella concept for this set of linked and overlapping features of reform design and implementation that flow from conceiving of reform as a one-off intervention. These were typically present in some combination in most reforms, although not all were always present. Reforms also differed in the extent of their projectization and in the specific ways that projectization manifested in their design and implementation.

Some reforms, such as Zambia's PSCAP (which we discussed in chapter 3), exhibited all these features of projectization. It was implemented by a dedicated project implementation team, was designed by donors and elite Zambian civil servants, then was rolled out across the wider civil service and had its own dedicated budget (funded mostly by donors but also by the government of Zambia) that had to be spent by the project's end date and accounted for against its predefined outputs. The three phases of PSCAP that were envisioned—developing and piloting new systems, rolling out of these systems to the whole civil service, and consolidation—even corresponded roughly to the three steps of Lewin's three-step episodic change model.[2]

On the other end of the spectrum, some reforms exhibited only a few of these features. For example, the basic structure of Kenya's Rapid Results Initiatives (RRI) was designed by a dedicated team, but the content of each individual initiative was codesigned with each specific ministry and aimed at restructuring core operational processes. The direct costs of the RRI and much of the staff time came from the line ministry rather than a project- or reform-specific budget line, and the intended operation of the RRI system in the civil service, in general, was indefinite. At the same time, the RRI system depended on a dedicated team of RRI coaches with its own funding stream, and each individual initiative was a short-term, one-off effort to make a specific change in each ministry. So as a reform, it still had some features of projectization.

Nearly all the reforms studied in this book had some, often most, of the features of projectization. However, measuring and classifying the degree of projectization exhibited by each reform is near impossible. To some extent, this is due to the practical challenges of inconsistent data availability across reforms. More conceptually, though, projectization is best understood not as a rigidly defined list of observable, binary characteristics but as a linked set of mechanisms that manifest differently across different contexts and types of reforms. How these mechanisms were related to the formal features of reform content and process was also highly variable. For example, even a seemingly binary feature like whether a reform has a predefined end date can be differentially reflective of projectization depending on its intent—for example, indicating an intent to stop reform activities in one situation but an intent to transition to a new phase of reform in another. There's also a sample selection problem: Some of the features that make a reform projectized, like an acronym and budget line, also mean that there is more likely to be documentation on it and that individuals are more likely to think of it as a reform, so it is more likely to be included in the reform histories I compile.

Subtle distinctions and challenges like this can be teased out and dealt with qualitatively for many reforms through the type of careful, triangulated description and analysis contained in the reform histories in the appendix and drawn on in these chapters. However, they would be impossible to consistently and precisely code in a quantitative fashion for each and every observed reform without oversimplifying the concept so much that it would lose most of its analytical force and without inadvertently conveying a false sense of precision. In chapters 3 and 4, I used some quantitative measures to summarize certain reform patterns (with many caveats); in this chapter, I'm not even going to attempt to quantify the extent or consequences of projectization. The rest of this chapter, therefore, analyzes these mechanisms of projectization qualitatively—albeit with as much precision as possible and closely grounded both in theory and empirical reform histories.

Let's now examine how the mechanisms of projectization manifested in the design and implementation of reforms in these six countries. I examine the features associated with projectization, how projectization led to reforms taking on those features, and their consequences for reform across five domains and submechanisms: expectations (i.e., goals), content, implementation, leadership, and politics.

HOW PROJECTIZATION SHAPED REFORM EXPECTATIONS

One of the most striking features of nearly all reforms was their extraordinary level of ambition. Reform plan documents almost always followed the same pattern. They began by deploring the existing state of the civil service and detailing its shortcomings, which they blamed on outdated structures, lack of motivation, and poor work culture. Then, they introduced a new reform agenda that promised to solve these problems, usually in a three- to five-year period. For example, an official pamphlet issued by Ghana's Office of the Head of the Civil Service described the Civil Service Performance Improvement Programme (CSPIP, 1994–2001) as intended "to improve efficiency and effectiveness in the delivery of services and outputs through: Institutional Capacity Strengthening in all Ministries, Departments and Agencies, Regional Coordinating Councils and District Assemblies; and Instituting a good governance culture in all aspects of the organisation and management of the Civil Service." It aimed "to do this through promoting performance improvement in individual institutions and by addressing efficiency, productivity, work ethic, service delivery, management and governance problems in the public sector." It then set out eight more specific objectives (with a further four added later) that were each also broadly defined, such as "i. promote the capability of civil service institutions to discharge their functions effectively in a transparent, competent and cost-effective manner thereby contributing positively to accelerated growth and equitable social development."[3]

Even in cases where the reforms' main motivation and content was oriented toward making fiscal savings or changes to salary levels, such as Ghana's Civil Service Reform Programme (1987–1993) and Single Spine Pay Policy (2007–2010), solving performance problems was cited as a major motivation and goal for reform.

There are several reasons why reforms tended to be couched in such ambitious language and expectations. One was the sheer magnitude of the shortcomings and necessary changes that reformers perceived. As Zambia's late Secretary to the Cabinet and Head of Civil Service Roland Msiska explained in an interview, "Anything less than overambitious in this country won't have a dent on our poverty."[4] Another was the difficulty in precisely measuring performance in a rigorous and comprehensive fashion across an entire civil service (as opposed to for a single process, service, or sector), which made it hard to demonstrate tangible improvement against a baseline and thus forced reform designers to resort to stating vague and overly broad goals that would be impossible to achieve. These were, perhaps, also compounded by the "planning fallacy"—the well-known behavioral bias of humans to be overoptimistic about how long tasks will take to achieve.[5]

Another major reason for overambitious goals, however, was the projectization of reform. Projects need to be approved in order to go ahead, and the existence of an approval process encourages reform designers to oversell the potential benefits of reform for the simple reason that the projects that appear most promising are the ones that tend to get approved. Anand Rajaram, former sector manager of public sector and governance work in Africa for the World Bank, explained that both governments and donors demanded ambition in projects:

> [Governments] are saying we have a big problem and if we say we are going to take a small crack at this then it is not inspiring. So part of it is built into the nature of that challenge that you have to excite the imagination by acknowledging the size of the problem in some way and that your effort will try to address it . . . saying we are going to solve the performance problem in four years may sound appropriately blood curdling and exciting but if no one has really bought into this then you have not spent the time building that openness and possibility. Then it is pure fiction what you have written. . . . I think that if you put together a project which says "we have very modest goals for this project, the reality of that country is that it is in dismal shape, but this will have a small effect on the conditions," why would the Board want to vote for that?[6]

Rajaram also pointed out that reform designers' personal incentives were often more closely linked to getting a project approved than to executing it successfully: "Unfortunately, the incentive system in a place like the World Bank is that you take the project to the Board and you are recognized. Whereas if you are the person who follows though, implementing the project and delivering it, there is much less structure." Kate Bridges and Michael Woolcock make a similar observation about projects being "overly ambitious" in their study of institutional reform in Malawi, citing World Bank research that found a "tendency to produce over ambitious plans at project design stage . . . for a [task team leader], what appears important when preparing the project is to make it as 'transformational' with a very ambitious agenda and please as many stakeholders as possible."[7]

The phenomenon of making projects overambitious to get the approval and resources necessary to undertake them is not restricted to civil service reforms or the operations of international donors but rather pertains to most large, complex public sector projects—and perhaps even to most major organizational change efforts in general. In his studies of infrastructure megaprojects, for example, Bent Flyvbjerg finds that nine in ten such projects have fewer benefits than forecasted and nine in ten overrun their anticipated costs. Examining the data, Flyvbjerg shows that this pattern isn't well explained by technical difficulties in forecasting or by simple overoptimism but instead that "planners and promoters deliberately misrepresent costs, benefits, and risks in order to increase the likelihood that it is their projects, and not those of their competition, that gain approval and funding."[8] Planners "spin scenarios of success and gloss over the potential for failure . . . this results in the pursuit of ventures that are unlikely to come in on budget or on time, or to deliver the promised benefits."[9] Similarly, Stefan Sveningsson and Nadja Sörgärde note in their textbook on change management that "To engage in organizational change can also be seen as an expression of drive and leadership, and can therefore enhance the status of those involved in it. Change attempts make it possible to profile oneself as a leader and to create an image of how you want to be perceived by others both within and outside the organization . . . hardly anyone wants to be seen as an ordinary supervisor or administrator of an existing organization compared to being seen as a change actor."[10]

Even within the context of civil service reform in Africa, setting overoptimistic reform goals was not solely attributable to donors. For instance, in one interview with Ghana's then-Head of Civil Service, Nana Agyekum-Dwamena, about the CSPIP reform, he explained how CSPIP had been adjudged to have fallen short of its goals because the project had included a number of highly ambitious "big-ticket items"—decentralization, public financial management reform, salary increases—among its deliverables. But the achievement of these was beyond the control of the Office of the Head of the Civil Service, and they did not happen within the project lifespan, so CSPIP was judged as having failed to deliver. I then began to ask a question about why the donors had been so overambitious in including these objectives in CSPIP, but Agyekum-Dwamena interrupted me: "No, no Martin, I disagree. In this particular case, with CSPIP, it was not the donors who set up those parameters." He then explained that these items were raised by domestic stakeholders during consultations in the design phase of CSPIP as being important for complementing and sustaining the performance improvement under CSPIP, and so including them within the CSPIP project had actually been the government's idea. "It was not an initiative of the DP [development partner], over a period of time it somehow then became a conditionality, not really imposed by DP, but by ourselves."[11]

Agyekum-Dwamena then carried on to explain that these system-wide, big-ticket changes were supposed to be undertaken under the umbrella of the broader National Institutional Renewal Program, which was governed by a National Oversight Committee. He said with a laugh, "My joke has always been that because they were called 'national oversight,' they lost sight of a lot of these things."[12]

To what extent did the projectization-induced incentive to set overambitious goals for reform explain the design and observed track record of civil service reforms? It's difficult to answer conclusively. The signs were all there—reform designers who talked about being ambitious and transformational and an abundance of reform plans that set out goals that were prima facie highly unlikely to be achievable. Issues of motivation, culture, and performance are always challenges for every bureaucracy in the world, so they are not problems that can be "solved" or "fixed," and certainly not in a few years. There is a lot of evidence that these dynamics exist with many types of public projects worldwide. At the same time, I encountered no "smoking gun" records of individuals saying that they consciously exaggerated the potential benefits or minimized the potential challenges of a specific reform, so it's hard to disentangle deliberate overselling from innocent overoptimism.

But the pattern of overambitious reform plans is certainly consistent with one of the more puzzling findings from chapter 3: that while most reforms achieved something, none of them achieved all the goals they set out for themselves. The dynamics of project design and approval processes are surely part of the explanation for this surprising pattern. Since it is nearly impossible to measure performance improvements civil service–wide in an absolute sense, the only available criterion for judging reform success is whether they meet their own goals. However, the goals, targets, and expectations for projects are endogenously determined, and reform designers have strong incentives to be overambitious in setting them—at least prior to the start of the reform. (Once the reform starts, of course, reform implementers' incentives might change, as the next section discusses.) So the more reformers think they can achieve, the higher they set their ambitions and vice versa. This practically guarantees that all reforms—from the most impactful to the least—will appear, from the outside, to be partial failures because they fell short of their own goals.

Another consequence of exaggerated expectations for reform is that it invites this glass-half-empty perspective, drawing attention toward reforms' shortcomings and away from what reforms do manage to achieve. Rajaram also reflected on this dynamic and what we should reasonably expect of reform efforts in complex bureaucratic systems:

> Even acknowledging that these things only achieve 50 percent of what they promised, that is fine. We should only expect 50 percent of what they promised to be achieved. I do not think they [can be] 100 percent with public sector reform, systems changes, behavioral changes, that would be difficult. Maybe only 20 percent even. . . . There are four components to this project, two went well and two did not. With the two that went well we met 60 percent or 70 percent [of] we thought we would achieve. That would not be a bad batting average for a public sector reform . . . [in] the evaluation of these projects, people do approach these projects like they are engineering projects. You say you [will] build 100 kilometers of road, but you only build five. You fall way short of the expectations. Building a road is something and building a system chain is quite different. It is way more complicated in some way . . . you won't get 100 percent [of] all these things [that] are there.[13]

HOW PROJECTIZATION SHAPED REFORM CONTENT

There is also reason to think that approaching reforms as one-off projects encouraged reformers to focus on making changes that were verifiable—in other words, on undertaking actions or outputs that could be specified in advance and measured afterward. Sometimes, this came in the form of creating or changing structures, rules, and processes, which exist on paper, and so are easy to explain in a strategic plan or annual report. Other times, it came in the form of undertaking countable actions, such as training sessions, organizational reviews, or purchasing tangible assets like computers. In this sense, the projectization of reforms compounded the tendency to focus on formal processes discussed in the last chapter, with similarly negative consequences for compliance with these processes and for eliciting important but unformalizable behaviors from civil servants.

The main factor driving this bias toward verifiable measures seems to have been the need to justify project budgets ex ante and account for how they were spent ex post. Getting either donors or finance ministries to approve budgets for unspecified reform activities is a challenge, and similarly, reporting on how resources were spent to donors or audit agencies is far easier for discrete, tangible outputs like new laws passed or workshops held than for harder-to-measure achievements like better implementation of existing processes or improved organizational culture.

These pressures have been widely noted by other researchers with respect to donor-funded institutional reform projects. For example, Lavagnon Ika has described an "accountability-for-results trap" in which implementers view results-based management tools as oriented toward external reporting and accountability rather than for use in internal improvement. Matt Andrews and colleagues have written about how projects tend to focus more on "form" rather than "function" in setting their targets and thus lead to changes in formal structures that are not matched by changes in actual behavior. Mark Buntaine and colleagues have shown that strong donor conditionalities push countries to adopt targets that measure "shallow" structural transformations rather than harder-to-achieve changes in outcomes. Kate Bridges and Woolcock have explored how donors' incentives to make disbursements on loans push them to adopt shallow targets in this mold.[14] In reflecting on Zambia's PSCAP-era reforms, Rajaram and colleagues ask, "is $10–20 million for a project in this area the best way to convince civil servants in a country to be more productive and change behaviors?" They go on to quote an anonymous Zambian senior official with experience designing and implementing reforms:

> In order to implement [reforms], I had been asked to provide cars to reforms teams, we did it; then, we were asked to provide computers, we did that too; then, we were asked to provide them formal training overseas, we did that as well; they came back and what happened? . . . Nothing! There was no greater capacity to reform despite these investments. Why is it so? Because reforming public sector requires a change in behavior and mindsets of people; cars, computers and formal training do not help in most cases. . . . The day a project is initiated, our problems begin.[15]

Most of this literature has placed the blame for this bias toward verifiable reform actions on the mindsets or incentives of donors. But similar dynamics also arose in cases where donors were not involved. For example, South Africa had relatively little donor funding for reforms (aside from technical assistance), but these tensions around the verifiability of reform achievements were very much present due to the threat of performance metrics being scrutinized or audited, and this sometimes pushed departments to focus these plans on activities and other easily measurable achievements rather than on outcomes. (Ironically, departments' five-year strategic plans were focused much more on impact and outcomes—because they were not subject to audit in the same way—and, thus, there was little articulation between them and the activity-focused annual plans.[16]) Similarly, South Africa had relatively few reforms with predefined end dates—perhaps due to not funding reform activities through donor projects—so, in that way, its reforms appeared to have been somewhat less projectized on average than in the other five countries. However, when I asked one senior official in South Africa whether he thought it was correct to say that the country had avoided taking a projectized approach to reform, he replied:

> I probably would disagree. . . . what we tend to do is we come up with new plans and we do tend to run it as projects. So unfortunately even a lot of our policies . . . what we tend to do is we start something and if we don't see immediate impacts we change it and give it a new name to try to create new energy around it. We tend not to show the patience to stick around and see it through. . . . My opinion is we have also very short-term focused projects and we don't see it through . . . we don't close these projects, we just slowly start putting them to the side and they become less and less visible.[17]

One specific way the bias toward verifiable reform activities manifested in both donor- and government-driven reforms was in the emphasis on trying to create new structures, rules, and processes rather than on trying to improve the implementation of existing ones. It is generally easier and more compelling (to most audiences) to say that a reform is going to revamp the country's annual appraisal system than to say that it is going to improve the execution of the existing one without making any formal changes. It is also easier to measure whether the new system has been formally adopted than how well it is being used. As one senior official in Ghana reflected, "Perhaps it's also the way we look at these reforms as projects. Projects come with money. Perhaps selling the idea that 'let's mend the old wineskin' is not so attractive."[18] For a whole range of stakeholders—from donors to finance ministries, audit agencies, voting publics, and even civil servants themselves—starting something new is easier to verify (and to claim credit for) than improving the operation of something that already exists. This bias in the content of reforms reinforced the existing tendency (discussed in the last chapter) toward formal reform measures and also had consequences for how projectized reforms were implemented.

HOW PROJECTIZATION SHAPED REFORM IMPLEMENTATION

The one-off nature of projectized reforms also directed energy and attention away from their effective implementation, institutionalization, and sustainability. Predefined timelines, the bias toward verifiable content, and a focus on introducing new processes rather than improving existing ones all contributed to far more attention being paid to making changes on paper than to embedding and sustaining them.

The senior official in Ghana mentioned above explained how this occurred in government. The combination of project timelines and pressure to disburse project funds meant that reformers spent most of their energy on the preparation and execution of the first phase of a reform initiative, often designing and introducing a new formal process. But by the time that was complete, either the funds and momentum had run out, or the project schedule stipulated moving on, and so the effective implementation of a reformed process was almost never the main focus of attention: "We have 3 phases, after we finish phase 1 instead of making sure this one is being used, we go onto phase 2." She blamed this implementation dynamic—rather than the content of the reforms, which she thought was largely good—for the pattern of disappointing reform impacts: "Probably we're not spending enough time cascading the ideas down, because as soon as the project ends it ends." She continued on to lament that individuals and governments tended to start one initiative, then move on to something new and forget about the older initiative instead of saying, "Wait a minute, let's make sure this one is being used before we go on."[19]

Similarly, Agyekum-Dwamena explained with respect to Ghana's Civil Service Reform Programme (1987–1993) how, as a junior officer working on the reform, he had "the impression later that there was going to be a review and the consultants are coming. 'We are to meet this deadline,' that was the message that was coming from OHCS [Office of the Head of Civil Service], MSD [Management Services Department], even our team leader that 'we needed to finish this thing by this date' so I did not really get the impression that we were doing it because it is good for the civil service." But when I asked a question about whether the donor-drivenness of the CSRP meant that its content was wrong for

Ghana, he replied, "No, no, no—I do not think it was bad, because all the things that we were supposed to do were good. But I think it was the timelines were more, we were doing because we have to deliver."[20]

This pattern of reform implementation—or nonimplementation—appeared across numerous sources, contexts, and reform efforts, both donor-supported and nondonor-supported. In Zambia, World Bank project completion reports from the Public Sector Management Program (2006–2012) state that service charters were "adopted and institutionalized" in eight ministries and pointed to the Ministry of Lands as a successful example.[21] But in an interview, a senior Zambian official who had been involved with supporting the adoption of service charters stated that he and his colleagues had observed that these tended to be adopted on paper and effective for one or two years but then rapidly drop off—even citing the Ministry of Lands as an example of this.[22] In Ghana, a different senior official observed a similar pattern of changes being put in place but then rapidly regressing in response to pressure brought to bear under the decidedly nondonor-driven Chief Directors Performance Agreement system since 2013.[23] An official in Nigeria described how SERVICOM evaluations of service delivery organizations were supposed to include periodic follow-up visits to check the implementation status of recommended changes but that this rarely happened.[24] In South Africa, another official explained that the raft of new laws and regulations introduced in the mid-1990s was accompanied by an initial burst of training and awareness raising, but, "I think it should have gone on for longer. Again, they assumed that the changes, I think they made assumptions about how, about the ease of implementing the changes, they made assumptions that things could be done in a relatively short period of time and that the initial capacity building that they did would be permanent. And they underestimated the extent to which it is difficult to implement changes like this and the length of time they take to implement properly."[25]

Why should it matter for implementation if a given process is introduced as a one-off intervention or project—particularly if such projects actually seem quite effective for designing and adopting new formal processes? Again, the answer goes back to the distinctions between nonperformance, perfunctory performance, and consummate performance from chapter 2. Formal processes are easy to adopt on paper, but if civil servants don't want to execute them—because they take time, create private costs, or perhaps are ideologically opposed to them—then they won't unless they expect that compliance will be enforced. So enforcement credibility matters for compliance. However, compliance alone only gets people to go through the motions, moving them and the organization from a state of nonperformance to perfunctory performance. For some types of processes, this might be adequate, but many important performance-related actions require individuals to do nonverifiable things that can't be formally prespecified and/or measured. Getting individuals to undertake these tasks as well—getting to consummate performance—requires that they expect other members of their team to also do so: bosses to recognize and reward it, and colleagues to match their effort so that it doesn't go to waste.

Expectations are about the future, about the continued emphasis on the implementation of these new processes. But approaching reforms as one-off, nonrepeated interventions, sometimes even with end dates, conveys the opposite message to workers who are the targets of reform: This is a one-time thing; it will not continue. Or at least, its effective implementation will not continue to be a priority for the people who authorize, lead, and manage reform efforts. Trying to achieve lasting changes with one-time interventions thus ignores the mechanisms that can lead to ongoing improvements in performance behaviors.

To illustrate this with a more tangible example, let's return to the case of individual-level performance management reforms that we discussed in chapter 4. Recall that these were intended to work via an annual process of setting performance targets, assessing performance, and delivering rewards or sanctions associated with that process. This is not a simple linear process where civil servants' actions follow in a direct and deterministic way once the system is introduced. Rather, it is a repeated cyclical process in which each step is shaped by expectations of how future steps of the process will unfold and by experiences of how the past steps have unfolded. Together, these experiences and expectations shape their perceptions of how credible the system is and how it will be implemented. As one Zambian official explained, civil servants' disappointing experience of previous reform waves meant that when new efforts came, they responded by saying, " 'Ah we've seen it before.' . . . It can make a very toxic environment."[26]

Many processes introduced as reforms are intended to be ongoing and repeated, like annual appraisal systems. In their review for the World Bank about the worldwide evidence on such systems in both the private and public sectors, researchers Sabina Schnell and colleagues blame the frequent ineffectiveness of these systems on how they are often imposed as one-off reforms: "having a performance management system is not enough. . . . Yet more often than not, organizations approach the introduction or overhaul of their performance management systems as a one-off change in human resource (HR) rules and procedures, rather than as part of a broader set of long-term reforms of various core organizational processes."[27]

Thinking about the implementation of these systems as a matter of changing repeated processes, as figure 5.1 illustrates, rather than as the rolling out of discrete

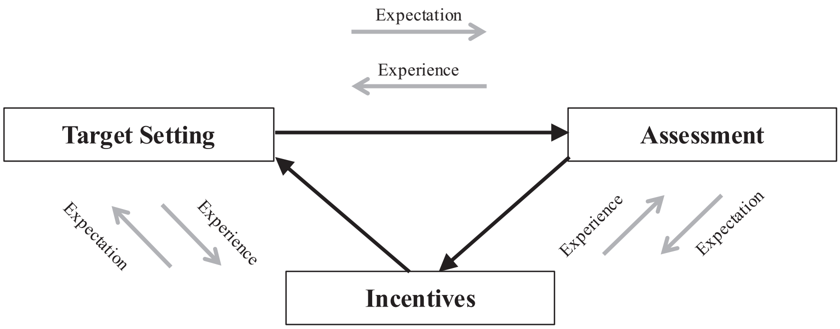

FIGURE 5.1 Performance Management Systems Are a Process, Not an Intervention

Source: Author's synthesis.

interventions helps explain why attempting to introduce them as part of one-off reforms often resulted in failure. Getting repeated processes to work properly requires not just creating them but ongoing efforts to support and reinforce their operation. But the dynamics of projectized reforms undermined reformers' incentives to do this and conveyed to the broader civil service that it probably wouldn't be done. Approaching the reform of ongoing processes as a discrete intervention thus undermined the effective implementation of such reforms.

HOW PROJECTIZATION SHAPED REFORM LEADERSHIP

Most reforms were characterized by top-down leadership models in which a central actor at the peak of the bureaucratic hierarchy designed and decided on the reform content and imposed it on the rest of the civil service. In some cases, this top-down reform approach was led by an individual like a president or head of civil service, in others by an organization like a cabinet office or ministry of public service, and often by some combination of the two. To some extent, this leadership model was identifiable by specific institutional and project features that were often present, such as the separation of reform from organizational processes and of design from implementation. But the dominant feature was an attitude held by those central actors who were primarily responsible for originating and delivering the reform about who was driving the reform and who "owned" it. This attitude was picked up and mirrored by rank-and-file civil servants, who generally perceived reform as something being done *to* them and their organizations rather than something being done *by* them.

This sentiment came through strongly in many interviews, especially with nonelite civil servants who were not involved in designing and implementing reforms but were affected by the rules, processes, and structures they introduced, many of whom specifically used the phrase "top-down" to describe how they experienced these reform efforts.[28] This pattern has also been noted by other researchers. For example, Abraham Rugo Muriu and Frank Kwaku Ohemeng both attribute the failure of individual-level performance management reforms to spur cultural change and meaningful performance improvement in Kenya and Ghana, respectively, to them being implemented in a top-down fashion and seen by civil servants as an imposition.[29] Similarly, Robert Dodoo (writing in 1997 as Head of Ghana's Civil Service), Stephen Adei and Yaw Boachie-Danquah, and Joseph Ayee all attribute the shortcomings of Ghana's Civil Service Reform Programme (CSRP, 1987–1993), in part, to the top-down fashion in which it was imposed on the civil service.[30] And with respect to institutional reforms in general, Matt Andrews has noted that donors tend to find it convenient to engage with a single reform "champion" within a government, leading to reforms often having only a very narrow support base.[31]

It is easy to see the top-down leadership model at work in reforms like Ghana's CSRP, which was oriented around cost-cutting, undertaken in conjunction with a structural adjustment program, and funded almost entirely by the World Bank. In such cases, reforms were clearly imposed on civil services by a combination of donors and a narrow set of elite actors within the government, such as project implementation teams and finance ministries.

However, the tendency toward top-down leadership of reforms was also in evidence in reforms that were definitely not donor-driven, such as the creation of Nigeria's SERVICOM, an agency that conducted service delivery reviews of other public sector organizations. Although it was inspired by a reform that happened in the United Kingdom and subsequently funded largely by a grant from the UK's Department for International Development, SERVICOM was created at the directive of President Olusegun Obasanjo and enthusiastically championed by him—supported by donors but not driven by them. But despite this high level of commitment from the presidency, SERVICOM initially struggled to be effective at spurring improvements in the ministries, departments, and agencies it worked with: "But at that time when the SERVICOM reform started, it was a top-down approach, [it] even started from the Presidency. He called the ministers and council and told them this is the directive, this is what you should do. And I think there was a little bit of . . . they felt they were being imposed

[on] . . . well, there was compliance. But [also] the undertone of [we] really don't understand what this is all about."[32]

Similarly, development consultant and former head of Nigeria's Bureau of Public Sector Reforms, Joe Abah, described how SERVICOM's designers focused on getting changes mandated from the top down in ways that ended up undermining their ability to get cooperation from the service delivery organizations it worked with:

> I think the effect of [SERVICOM] has been limited. It hasn't been a complete waste of time but the effect has been limited . . . there's this mentality of "all we need is power." That's the first thing they did, was go to the Federal Executive Council to ask that everybody must have a SERVICOM office. So of course everybody set up a SERVICOM office . . . people stuck the SERVICOM banner on the nearest toilet, and nobody cared what was going on inside so it was just appearing to comply, which they did. The next thing they did was [say], "oh the nodal officer should report straight to the minister, period," so here is the permanent secretary rubbing his hands and thinking, "ok so you're a nodal officer, you are an assistant director, you're going to bypass me the permanent secretary to go report to the minister and you expect me to release funds for you to do any work . . . never going to happen." So again, it antagonized the system against itself. . . . Fine. Have a SERVICOM office [and] post to the most problematic person in the office in that place to get him out of the way. Make sure you don't release any funds to them.[33]

But despite these challenges, there is also evidence that SERVICOM did manage to get many organizations to implement some positive changes that improved service delivery, even if not to the extent that it desired. Support from above—both from the presidency and senior managers within each organization—was crucial to the setting up of SERVICOM and much of what it was able to achieve.[34] But it also illustrated the limitations of heavily top-down approaches for achieving and sustaining performance improvement.

Interestingly, SERVICOM has gradually evolved away from this top-down approach. After the exit of President Obasanjo in 2007 and the end of DFID funding in 2009, "Everything basically ground to a halt,"[35] and for half a dozen years, the organization struggled to sustain itself, retain staff, and continue its core work. However, the organization managed to cobble together funding from the government budget and international donors in order to maintain itself and operate on a more sustainable footing. While it arguably did not regain the political salience and centrality that it had during the initial Obasanjo years, it adjusted to this, in part, by pivoting its operations increasingly toward universities and hospitals rather than the higher-profile and more powerful (and, hence, potentially resistant) government ministries and agencies on which much of its early effort had been focused.[36] It also launched a weekly SERVICOM Help Desk radio program, in which SERVICOM's National Coordinator would listen to complaints about service delivery and human rights issues from callers, give advice, and follow up on cases with the relevant institutions—combining raising public awareness with generating legitimacy for its mission beyond the highest echelons of government. In this way, SERVICOM also serves as a model of what can be creatively achieved by working outside of a purely top-down paradigm.

HOW PROJECTIZATION SHAPED REFORM POLITICS

A truism among practitioners and researchers alike is that reforms require political backing and ownership at the highest levels to be successfully implemented. Numerous case studies and project reports have lamented the lack of political support for reforms and sought to explain the causes and consequences. The ways political support can be helpful for reform are obvious, and among the reforms examined in this book, there were certainly many whose shortcomings were, at least in part, attributable to a lack of political backing.

But even for reform efforts that did receive strong political backing or originated from politicians, this close association sometimes proved to be a double-edged sword. The same mechanisms that allowed politicians to channel their support into greater clout, attention, and resources for reforms also led them to be designed and implemented in more projectized ways that undermined their implementation, impact, and sustainability.

Where it occurred, the tight association of reforms with political leaders and parties seems to have arisen from a combination of two factors. First, bureaucrats, donors, and politicians alike often perceived vocal political support for a reform as necessary for its success and, thus, sought to create and channel it to support reform efforts. Second, politicians often sought to emphasize their association with a reform to claim credit with either voters or donors.

Both of these factors pushed all actors involved to find ways to make reforms more closely associated with their political sponsors. One obvious way was to separate reform activities from core bureaucratic processes by denoting reforms with their own acronyms, branding, and budget lines, all of which made it easier for politicians to attract attention to these reforms, make them more visible, and claim credit for them. Another way was to drive reform design and implementation through separate project teams or reform implementation units—often located in the presidency or otherwise close to executive authority—rather than through the mainstream bureaucracy, which gave politicians the ability to closely oversee reform activities and gave reform implementers greater access to their political sponsors. The association of reforms with political leaders strengthened the tendency toward both of these features of projectization (although they were not the sole cause of them). These same factors also reinforced the other submechanisms of projectization discussed above—more ambitious-sounding goals and expectations, more verifiable and publicly visible content, more one-off implementation, and more top-down leadership.

One of the ways these dynamics may have manifested was through a shortening of reform time horizons, as electoral pressures pushed leaders to not only want to demonstrate fast results but also to become risk-averse near elections. For example, Zambia's former Secretary to the Cabinet Roland Msiska lamented in an interview that reformers only had three years of any electoral cycle to work in because during the last two years, politicians would be campaigning and any complaint against a reform would sink it.[37] Similarly, a South African civil servant with experience implementing reforms reflected: "We've got to appreciate that politicians do not think longer than five years in advance. We've got elections every five years, so they are pressurized to show good, fast results. And then if you're not getting results from a specific initiative, then you can close it or just let it dwindle away."[38]

The incentive for politicians to create a perceived identification between themselves and specific reform efforts may also have undermined the likelihood that these efforts would be sustained after they left. For example, one interviewee involved in reforms in Ghana complained that politicians each wanted to start their own reforms once they came into office rather than continuing existing ones started by previous leaders.[39] This is consistent with the idea that political leaders' support for reforms is driven, at least in part, by a desire to claim credit for these initiatives. The more publicly a reform is identified with a given politician, the more likely their successors would likely be to discontinue it. Similarly, the researcher and experienced reform architect Tunji Olaopa remarked, "The backlash created by personalization of reform by [senior members of President Obasanjo's administration] . . . they made some statements at the time that offended the civil service. You know, 'that service has no brain,' 'most of them are archaic' . . . and this thing filtered into the press. . . . It eroded the service['s] support to some of the reforms that they did, so consequently when they exited the service was more inclined to pull down [some of] what they are doing [rather than] to drive it forward."[40]

How common is it for political backing to intensify the projectization of a reform, and how negative are the consequences for the reform's success? The answers are unclear. On the one hand, it is logical that greater personal identification of a reform with a political leader would lead to an increase in some of the features of projectization, and there is ample evidence consistent with these mechanisms across many reform efforts. On the other hand, the analysis in chapter 3 found little support for the idea that electoral cycles and leadership transitions were the main factors driving reform initiation and duration. It seems most appropriate to conclude that, while political time horizons do shape reform in at least some cases, they are not necessarily the main factor doing so—just as political incentives are only one of several channels through which projectization emerges and shapes reform efforts.

Similarly, while it seems plausible that politically driven incentives for projectization accentuate some of the negative consequences of projectization outlined above, it does not necessarily follow that greater political support is harmful overall for reform efforts. Rather, it points to a need to think in greater detail about the different ways political support can be expressed publicly and channeled bureaucratically during reform efforts. The challenge for reform leadership is to maximize the benefits of political support while minimizing the potential negative consequences of greater projectization, all while preserving politicians' incentives to offer support in the first place. We will return to this issue in chapter 7 when considering potential alternative approaches to reform that avoid the pitfalls of projectization.

• • •

This chapter has argued that governments' reform efforts were usually characterized by projectization—an approach that sees reform primarily as a one-off intervention that is separate from core processes, has distinct design and implementation phases, and/or has an envisioned end date. Projectization shaped reform and undermined its effectiveness through five sets of mutually reinforcing submechanisms that related to the expectations, content, implementation, leadership, and politics of reform. These patterns of projectization thus constituted the second main mechanism of reform failure I observed, and are summarized in table 5.1.

TABLE 5.1 How projectization shaped reform efforts

	Common features of projectization	Why?	With what consequences?
Expectations	• Overly ambitious goals • Framed as solutions to wide range of performance problems	• Need to exaggerate expected benefits to get buy-in and resources to undertake reform	• Reforms inevitably fail to achieve their goals, get viewed as failures
Content	• Focus on making changes that are ex ante specifiable and ex post measurable (i.e., verifiable)	• Design of project prior to and separate from implementation • Need to account for how project resources are spent	• Reforms focus on making formal changes rather than changing behavior • Reforms focus on introducing new processes rather than improving existing ones
Implementation	• Little focus on embedding and effectively implementing rules and processes	• Introducing new processes can be done as a one-off project or intervention; embedding them takes time and iteration	• One-off projects fail to improve compliance with formal processes or encourage informal supporting actions • Changes on paper but not in practice
Leadership	• Reforms designed and delivered in top-down fashion by a central actor	• Civil services conceived of as systems in bad equilibrium in need of a shock from outside/above • Separation of reform activities and teams from core organizational processes	• Reforms fail to get broad buy-in or trigger decentralized action, improvement
Politics	• Reforms closely associated with a specific political or bureaucratic leader • Greater separation of reform branding and activities from core bureaucratic processes	• Perceived need for high-level support and backing for reform • Desire to claim credit for reform (either vis-à-vis donors or voters)	• Higher salience makes it easier to push through formal changes but also decreases likelihood they are sustained after leadership transition • Shortening of time horizon of reform

Source: Author's synthesis.

Some clarifications and caveats are in order. First, projectization is not a single feature or variable but a linked set of mechanisms or submechanisms that frequently co-occur and stem from the same underlying factor: approaching reform as a one-off intervention. However, these mechanisms manifest to different extents and in different ways across different reform efforts, not only in the formal design features of the reform and its implementation structures but also in the mental models of the individuals driving the reform. So whether a reform has the features of projectization is not a simple yes/no question or even a linear spectrum but, rather, a hard-to-measure and interrelated set of qualitative judgments about how the reform is being approached. While the concept resists oversimplification, it is also abundantly evident in the data that essentially all reform efforts had some of the features of projectization, and many had all of them—with predictable consequences for how these reforms unfolded and consistently fell short of achieving the changes they aimed for.

Second, at the same time as projectization was clearly a mechanism of failure, it is not possible to conclude that projectization is always detrimental to reforms. Some of this is due to the empirical and methodological problems of measurement and the unavailability of counterfactuals that we've discussed previously, which make it impossible to conclusively estimate the impact of particular reforms. While we can observe reform features and approaches and trace through how they appear to be contributing to the failure and success of the reform, we need to be cautious about extrapolating from these mechanisms to the overall impact of the reform. More theoretically, it's also possible that projectization (or at least some features of it) is appropriate for some types of reforms. In particular, if a reform is trying to make a change that is highly verifiable and has predictable and deterministic consequences on performance—like flicking a light switch on or off—then approaching the reform as a one-off intervention might make sense.

However, what this book has found, over and over again, is that performance-oriented reforms in civil services are not like flicking a light switch, pulling a lever, or changing a line of computer code—they don't automatically and instantly translate into changes in behavior. This is because so much of what civil servants do is nonverifiable, and getting them to do nonverifiable things requires them to believe that their actions will be recognized and reciprocated, not just in the present but also in the future. This poses an inherent challenge for reform efforts that approach making change solely as a one-time intervention. So while some elements of projectized approaches might be appropriate or practical even in the most successful reforms, it's important to recognize the ways they risk undermining the end goal of reform: getting individuals to do the informalizable things they need to do to get to consummate performance.

It seems appropriate to close this chapter on projectization with a brief digression on the history of the term. The first use of the word *projectization* with respect to civil service or institutional reform that I have found comes not from an academic theorist but from the World Bank itself in a 1994 discussion paper entitled, "Projectizing the Governance Approach to Civil Service Reform," which was intended as a how-to guide for officials trying to package institutional reforms in ways that could be supported by project loans.[41] So from the term's beginning, *projectization* was understood as an effort to simplify and bound what was acknowledged to be the messy reality of systemic bureaucratic reform into something that was easier to administer and manage. Other practitioners and researchers have since remarked on what they term "projectitis," or the tendency to try (and fail) to achieve systems change through isolated projects with their own external funding streams in everything from health systems reform to solar system installation.[42] Also in the early- to mid-1990s, terms like *projectization* and *projectification* began being used in the literature on project management and IT investment in private firms in high-income countries—entirely separately from issues of development, donors, and public sector institutions—with *projectization* referring to "the extent to which a business is based on projects and the project way of working pervades."[43] The breadth of the term's usage is just one more piece of evidence that, while donors' involvement may exacerbate projectization, it is far from the sole factor driving it. Indeed, the negative consequences of projectized approaches to reform are also evident in much research on reforms, even in high-income countries where donors are absent. For instance, in their review of reform efforts in the United Kingdom, researchers Charlotte Pickles and James Sweetland write, "even where good ideas for reform are instigated and a serious implementation plan is in place, embedding and sustaining those reforms requires a different set of actions. Whitehall reform cannot be seen as a 'once and done' process."[44]

Having examined the two main mechanisms of reform failure, the next chapter turns to the main mechanisms of success that characterized reform histories across these six countries. While most reforms were designed to have an impact through the one-off introduction of new formal rules, structures, and processes, the most consistent benefits for performance occurred through a more informal and often unintended channel: getting civil servants to talk to one another, share information and ideas, and learn how to do their jobs better. Let's now explore how these unformalizable conversations unfolded, often in the shadow of (and sometimes despite) formal processes, and consider what lessons this yields for designing and implementing reforms.

6

Mechanisms of Success

Opportunities and Energy for Performance Improvement

Chapters 4 and 5 focused on the mechanisms through which reforms failed: trying to force performance improvements through formal processes and approaching reform as a one-off project or intervention. These mechanisms meant that reform efforts tended to yield perfunctory compliance with the new processes—just going through the motions—rather than meaningful changes in behaviors and performance. And even the perfunctory compliance often faded away over time.

However, many reform efforts also succeeded in spurring genuine improvements. In varying ways and to differing extents, they got civil servants to undertake unformalizable actions, changed expectations and norms, and made performance more prominent in organizational cultures. These weren't magic bullets that completely transformed civil services overnight. They were usually limited in scope, sometimes achieved their positive effects through unintended or unforeseen pathways, and had their own shortcomings and limitations. Still, they were promising examples of how system-level reforms can lead to meaningful change despite all the challenges they face in doing so. What were the mechanisms of success that explained how such improvements occurred?

This chapter argues that when reforms succeeded at improving performance, it was because they created *opportunities* and *energy* for civil servants to talk about performance and how to improve it.

Opportunities came from a mix of formal processes and informal spaces that led civil servants to have conversations about goals, roles, and performance that were not happening before. Providing these opportunities helped empower the many civil servants who already cared about performance and wanted to improve and gave them outlets through which to engage with their colleagues. This contrasted with the dominant paradigm of assuming that carrots and sticks were needed to force behavior change on unwilling civil servants. These forums and processes could help make performance and results more salient within organizations, and they also served as opportunities for civil servants to learn about their roles and how they connected to those of their colleagues, to give and receive feedback, and to generate and share ideas for improvement.

Energy meant making civil servants believe that the reform effort would actually lead to something tangible and wasn't just talk that would soon dissipate. It helped reformers surmount the credibility barrier that faced reforms, in which individuals did not change their behavior because they didn't expect their colleagues to change anything either. Energy came from leaders and was crucial for focusing attention on performance and building credibility and, thus, momentum for widespread changes in both nonverifiable and verifiable behaviors.

Even in the absence of energy from leaders, the intrinsic motivation and professionalism of individual civil servants still sometimes drove change at smaller scales when opportunities had been created. Similarly, energy for change in the absence of structured opportunities for discussing performance was also sometimes enough to spur improvements. But when reforms did succeed in making positive changes at scale, it was usually because these two mechanisms combined to catalyze conversations that helped managers and rank-and-file civil servants do their jobs better on a day-to-day basis and that contributed in small but meaningful ways to changing organizational cultures.

Before proceeding, an important reminder: In this book, I focus on analyzing not *cases* of success and failure but *mechanisms* of success and failure. Rather than classifying reforms as "successes" or "failures" and then trying to explain this variation, my main aim is to identify and understand the mechanisms through which each reform succeeded in some ways and failed in others. To recap, there are two reasons for this. First, the criteria for assessing reforms—their goals or targets—are set endogenously in the reform design process, so it is difficult to accurately measure and compare overall reform success. The fact that essentially all reforms are partial successes (or partial failures, depending on your view) is partially due to this. We can use qualitative data and careful triangulation to understand degrees of success and failure, but it would be misleading to try to formally code it. Second, and more importantly for this chapter, each reform is not wholly characterized by either mechanisms of success or mechanisms of failure but by multiple mechanisms coexisting together in various combinations and interacting. The reforms I discuss as exhibiting these mechanisms of success are not necessarily "success stories"; indeed, in past chapters, I have analyzed some of these same reforms in terms of their mechanisms of failure. So from an analytical perspective, the question is not so much whether each mechanism is present but to what extent and in what ways.

To investigate these mechanisms of success, this chapter begins by returning to the individual-level performance-linked incentive reforms discussed in chapter 4. It shows how they sometimes succeeded in creating opportunities for discussing and improving performance despite the near-universal absence of the rewards and sanctions that were intended to accompany them. The chapter then discusses how other types of reforms were also sometimes able to create opportunities for discussing and changing practices. It then turns to examining a set of reforms that succeeded in generating energy for change but created few opportunities for civil servants to actually discuss how to do it. Finally, the chapter considers how reforms did (and did not) manage to combine these two mechanisms of success—opportunities and energy—in various ways. The chapter closes by discussing some of the challenges of activating and sustaining these mechanisms and the ways projectization and a focus on purely formal changes often undermined them. Taken together, this chapter sets the stage for the third and final part of the book, which seeks to understand how reformers can maximize mechanisms of success and minimize mechanisms of failure.

CREATING OPPORTUNITIES TO DISCUSS PERFORMANCE

In chapter 4, we examined how individual-level performance management systems, like annual staff appraisals or leadership performance contracts, consistently failed to deliver differentiated rewards and sanctions. These systems' focus on using formal structures and processes to try to force or incentivize better performance constituted one of the main mechanisms of failure for reforms.

Despite the dismal record of formal incentives, however, it is not the case that individual-level performance management reforms had no positive effects. Across countries, interviewees pointed to the positive benefits that such systems sometimes brought by creating discussions around goals, responsibilities, and performance. For example, one rank-and-file civil servant in Ghana explained an effort in the mid-2010s to revitalize the annual performance appraisal system:

> The system gives everyone a specific focus and I benefit from that. I know what I am expected to do by the day, month and year. It makes me stay focused and more efficient. It makes me want to deliver and gone are the days where I sit about waiting for the work to come. . . . Before, there was no mutuality or

> participation. . . . Now you sit with your supervisor. . . . Before you would set [targets] and there is no assessment until the end of the year . . . there is no way to see what has happened, [but] now there is a way to review. The old annual appraisal was sparingly used, only when people were due for promotion. Now there is care behind it. [Formerly the] head of department that would do [the assessment], now it is the one you are working with who will do the appraisal.[1]

Similarly, another midranking civil servant in Zambia said of their annual performance appraisals: "There has been a one-on-one interaction which has helped with understanding what gaps people have. It has helped me to understand at what level they are supposed to operate because at the end of the year, we find out if that has been met."[2] These discussions helped civil servants better understand how their work fit into that of their team and the broader organization. As one rank-and-file civil servant in Ghana commented about the country's Chief Director Performance Agreement system in the mid- to late-2010s (which had also begun to be cascaded down to directors), "You realise what you do is indirectly linked to the director. If I deliver, then the chief director is able to satisfy what he has to do. If I do not then my director and chief director are affected."[3] And a 2008 Zambian consultancy report evaluating Zambia's staff performance management systems noted, "The performance evaluation systems that both managers and employees rate highest are those that let managers and employees communicate—share ideas, opinions, and information whereas most 'traditional' systems put managers into the position of uncomfortable judges, telling employees that their work either was or was not 'satisfactory.' "[4]

These conversations and the clarity and understanding they brought were widely reported to have been useful, even though they did not always occur in each reform case or uniformly across organizations and teams. So while performance management systems were generally introduced with the idea that attaching incentives to formal performance measures was the key to improving performance, the main mechanism through which these appraisal systems actually improved performance was through the unformalizable processes of discussion and feedback, and civil servants then choosing to use this information to take largely informal, nonverifiable actions to improve performance. These discussions were largely nonverifiable—hard to prespecify exactly ex ante, hard to measure well ex post—because, while the formal rules of a performance management system could demand that such a meeting happen, they couldn't force the discussions to be honest and meaningful rather than perfunctory. So if the main mechanism of failure of performance-linked incentive systems was their focus on formal processes and formal performance measures, their main mechanism of success was their catalyzing of informal discussion and communication.

Of course, creating clarity around targets and feedback on performance was an explicit goal of many individual-level performance-linked incentive reforms. But in practice, this goal tended to sit in tension with these reforms' primary objective of creating a system of objective targets and performance measurements based on which incentives could be delivered. This was because trying to achieve full objectivity required squeezing out all the nonverifiable aspects of target-setting and performance measurement and feedback that inherently created ambiguity, subjectivity, and potential unfairness.

For example, individual appraisal and performance contracting systems universally demanded that targets be S.M.A.R.T.: specific, measurable, achievable (or attainable), relevant (or realistic), and time-bound. In other words, they should be verifiable: perfectly specifiable in advance and perfectly measurable after the fact. But this focus on measuring performance through verifiable targets fits poorly with the reality of civil service work. One South African civil servant involved in numerous reforms reflected on this mismatch:

> We try to be S.M.A.R.T., but I cringe every time I get feedback from my HR unit telling me my indicators are not S.M.A.R.T. A lot of people try to make performance management an objective system, and I tell them it cannot be . . . for me [discussion] is the most critical thing in this whole performance management system . . . [having] regular information about what's happening, and then having regular feedbacks. And you have to acknowledge that it's a subjective thing. Yes you can have some objective measures, but they should be to substantiate your subjective opinion as a manager and point towards indicators of that.[5]

This tension became more severe when stronger rewards, sanctions, and controls were applied to these processes. For example, in Kenya, linking hard incentives to performance contracts reportedly pushed permanent secretaries to start to seek out soft targets that they could control rather than aiming at improvement in final outcomes.[6] And in South Africa, the emphasis on measuring outcomes and results reportedly led not just to civil servants devoting a huge amount of time to an increasingly burdensome performance reporting system but also to performance agreements and annual work plans that increasingly focused on highly measurable short-term deliverables that were largely delinked from the broader five-year national plan—the implementation of which was the original rationale for the focus on outcomes.[7] Thus, while formal performance appraisal and performance contract systems had the potential to create discussion opportunities that could help improve performance via better communication, this benefit was easily erased when high-powered incentives were attached to these systems.

Aside from individual-level performance management systems, many other types of reforms also created opportunities for discussing performance and how to improve it. In South Africa, for example, the annual Senior Management Service conferences were "excellent" forums for managers to communicate with one another, share ideas, and shape culture.[8] "The speakers we brought to those, the exposure we had, it really benefitted [us] . . . talking to a number of senior managers, they seem to have really benefitted from that exposure that made a huge difference."[9] In Nigeria, the SERVICOM service delivery improvement agency conducted in-depth, participatory service delivery evaluations with scores of ministries, departments, and agencies to identify problems and ways to improve.[10] In Senegal, the Comité d'allègement et de simplification des formalités et procédures administratives (Committee on Reducing and Simplifying Rules and Administrative Procedures) regularly convened a rotating and cross-institutional set of officers to identify opportunities to alleviate both internal and client-facing administrative burdens.[11] In Kenya, Rapid Results Initiatives (RRIs)—one-hundred-day periods of intense action in a ministry aimed at changing a specific process or attaining a specific objective, with coaching (but generally not financial resources) from a central support team—brought organization members together to collectively diagnose and find ways to fix longstanding problems.[12]

What united these diverse reforms was the thoughtful use of formal processes to convene, enable, and encourage discussions about performance. The processes themselves were mainly formal in the sense that they could be specified and measured, but the discussions themselves were largely unformalizable. For example, annual staff appraisal systems could stipulate that supervisors and supervisees should have annual or quarterly meetings and talk about performance, and the completed appraisal template could be collected as proof that the meeting occurred (although even this could be faked, of course). But no one could force the content of these conversations to be meaningful, and efforts to do so via stronger incentives and control mechanisms tended to lead instead to gaming behavior and resistance.

Instead, the motivations that led civil servants to take advantage of these opportunities for discussion were intrinsic ones: the desire to do better for the country and the organization's clients and the professionalism that led people to want to get things done and feel proud of their teams and organizations. Of course, this meant that the take-up of these opportunities for performance discussion was far from the universal, homogenous, mechanistic ideal that many people envisioned for these systems and processes. Many supervisors and supervisees treated appraisal discussions as a formality or did not even have them, and many organizations treated SERVICOM evaluations as an imposition to be endured rather than a chance to improve, and so on. One interviewee remarked about South Africa's Management Performance Assessment Tool, which conducted annual reviews of organizational management practices and sought to use these to spark conversations around improvement, "It was more like a support measure than a regulatory measure, MPAT. And a support measure, support can be offered. But you can take a horse to water but you can't force it to drink. That was always going to be a limitation with it. And I think it was useful for those who wanted to participate and wanted to improve the administration across the board. It was useful for them."[13]

So creating these opportunities for discussion didn't automatically lead all people or organizations to do nonverifiable things. At the same time, neither did the use of rewards, sanctions, or compliance tools. But at least in some cases, the creation of these opportunities was a successful mechanism for getting civil servants to undertake important but nonverifiable tasks.

CREATING ENERGY TO DISCUSS PERFORMANCE

The other main mechanism through which reforms got people to change unformalizable behaviors was by creating energy around improving performance. More precisely, reforms sometimes managed to shift individual and collective norms and expectations about acceptable or encouraged behavior, how other colleagues were likely to behave, and whether unformalizable actions would be recognized and rewarded. A major common barrier to reforms was the sense of inertia and skepticism that often greeted them, both due to cynicism induced by past experience and because even the most motivated individuals would only want to put out the extra effort to perform consummately if they thought that other members of their team and organization might do the same. Creating energy in this way could help break out of this negative equilibrium, activating individuals' intrinsic motivation and desire for better collective performance and helping make them believe that taking important but unformalizable actions would not go to waste.

Perhaps the clearest example of this was South Africa's Batho Pele ("People First") initiative. Batho Pele was launched in 1997 as part of a postdemocratization raft of legislative and policy reforms that aimed at transforming the apartheid-era bureaucracy, which had been oriented around controlling and oppressing the majority population rather than being responsive to its needs and wellbeing. The white paper that inaugurated Batho Pele called it: "an approach which puts pressure on systems, procedures, attitudes and behaviour within the Public Service and reorients them in the customer's favour, an approach which puts the people first. This does not mean introducing more rules and centralised processes or micro-managing service delivery activities. Rather, it involves creating a framework for the delivery of public services which treats citizens more like customers and enables the citizens to hold public servants to account for the service they receive. A framework which frees up the energy and commitment of public servants to introduce more customer-focused ways of working."[14]

The Batho Pele approach was broken down into a set of eight principles, and it was these principles rather than a set of activities or outputs that comprised the core of the reform: "Batho Pele is not a single project. . . . Batho Pele is a characterisation of the nature and quality of service delivery interface that should obtain between government and the public."[15] The slogan rapidly became widespread throughout the civil service through consistent repetition by two successive Ministers of Public Service and Administration, Zola Skweyiya and Geraldine Fraser-Moleketi. Although the Department for Public Service and Administration did undertake a set of awareness-raising activities and developed some implementation guidelines and requirements—I'll return to these below—Batho Pele was mainly driven by leaders' rhetorical efforts to create energy around reorienting and improving the civil service's culture. One interviewee commented that "it was very much up to departments to implement it themselves. Obviously some departments saw the value in it and used it to guide the way they do things, but for other departments it was at best a compliance thing that we had to do and submit reports."[16]

Judging the impacts and success of Batho Pele as a reform is difficult due to its broad scope and ambition, its relatively decentralized implementation model, and its focus on difficult-to-measure culture change rather than the delivery of specific outputs. But there is general consensus that the Batho Pele slogan and rhetoric became widely recognized throughout the civil service down to frontline officers and even members of the public.[17] There is also consensus that some departments, such as the Department of Home Affairs, dramatically transformed their processes and client orientation through reforms under the Batho Pele banner. However, these successes in some areas were matched by lagging performance in others, both across departments and within them. For example, a set of service user surveys conducted by the PSC in 2012 found that majorities of Department of Home Affairs service users indicated satisfaction with the Department's levels of courtesy, information provision, and publication of service standards but that majorities also expressed being unaware of the Department's efforts on several other dimensions of Batho Pele standards.[18] But while its impact inevitably fell short of its hugely ambitious goals, as a reform, it seems to have been fairly successful in transforming deeply engrained norms around responsiveness to clients and service delivery—albeit, perhaps more widely so in principle than in practice—and in empowering civil servants with the motivation to make process improvements to do so.

Of course, part of what made Batho Pele successful in creating energy for change was not just the reform itself but also the moment of political and social transformation in which it was designed and implemented. Such moments sometimes helped reforms create energy for change within the walls of civil service institutions by linking them to wider changes and energy in society. For example, the most vigorous and successful civil service reforms in Kenya's modern history occurred in the years following the election of President Mwai Kibaki in 2002, which brought the country's first change in political leadership in decades and was widely seen as a clean break from the previous era of stagnation and economic crisis. This was as true for the civil service as for the country as a whole, with experts and former officials using terms like "euphoric" and "energized" to describe the enthusiasm within the civil service at this time.[19] This energy was channeled into change in policy and processes through reforms such as the Economic Recovery Strategy, Results for Kenya program, a Performance Contracting system for permanent secretaries and their ministries, and the aforementioned Rapid Results Initiatives.

While reforms were sometimes aided in creating energy by piggybacking off wider social and political momentum, energy from outside the civil service was neither sufficient nor necessary for creating energy within the service. For example, the coming to power of Abdoulaye Wade in 2000 in Senegal's first postindependence "alternance" (i.e., change in party control) created a structurally similar moment to Kibaki's election in Kenya but created relatively little new energy for reforms to the civil service—although a number of good governance reforms were implemented in other parts of the government, and civil servants themselves were fairly successful in carrying over and rebranding much of the reform activity from the previous administration. Similarly, Ghana's CSPIP reforms in the mid- to late-1990s were partially successful in creating energy and a new cultural focus on performance and service delivery despite being initiated and executed during a period not marked by major political change. Instead, reform energy came from the new Head of Civil Service, Robert Dodoo, and his team at the Office of Head of Civil Service, developing a "homegrown" reform agenda that empowered and supported organizations across the service to come up with their own ideas and initiatives for improvement. So energy for reform didn't necessarily need to come from a president or other political leader; it could also be built within the civil service.

The idea that creating energy for change is a good thing may sound simple and obvious, but these and other cases illustrate just how difficult it was for reforms to achieve it. For every instance in which genuine energy for reform spread throughout a civil service, there are many more where leaders declared the need for change—often using nearly identical rhetoric—but failed to shift civil servants' expectations about whether this was just another set of superficial changes that would fail to translate into real practices as opposed to a meaningful break with the past. Even when political leaders had energy for reform and wanted to convey it to civil servants, it was easy to miscalculate how such messages would land. The derogatory comments about civil servants made by some senior members of President Obasanjo's administration in Nigeria, for example, reportedly created resistance to reforms among civil servants at a moment when democratization and broader social change should have helped build momentum for reform.[20] Creating energy for change was not just a simple matter of top-down rhetoric but a more nuanced process of articulating a vision, building credibility around it, and creating space in that vision for civil servants to imagine themselves as agents of change rather than objects of it.

COMBINING OPPORTUNITIES AND ENERGY

There are obviously potential complementarities between these two mechanisms of success, ways in which the mechanisms are more effective in combination than on their own. Opportunities for discussing performance will be more likely to be seized and to translated into action if civil servants believe that things are really going to improve and that their leaders and colleagues also believe this. Likewise, creating energy for improving performance will be more effective if there are meaningful opportunities for civil servants to channel it into discussion of how to improve and tangible actions. How successful were reforms at bringing these two mechanisms together?

There are certainly some examples in which these two mechanisms appear to have complemented each other. In Kenya, for instance, the broader political and social energy for reform after 2002 was channeled into the civil service through (among other reforms) a Performance Contracting system that defined organizational goals, made them salient to leaders and staff of the organization alike, measured performance, and (for a few years) gave a group bonus to members of the top-performing ministry. This new salience of performance meant that when the Rapid Results Initiatives coaches turned up to offer their services, organizational leadership was more eager to accept their assistance and put effort into the one-hundred-day improvement initiatives that were framed in terms of the goals of the performance contracts. More broadly, many organizational performance review schemes brought together energy and opportunities to diagnose problems and identify solutions in participatory ways, including Ghana's beneficiary surveys, self-appraisal instruments, and performance improvement plans under CSPIP; South Africa's Management Performance Assessment Tool, which sought to evaluate not only if each management process was being formally complied with but also if it was actually being used to improve performance; and Nigeria's SERVICOM institutional diagnostics, which often received more welcoming receptions because of the high-level political backing for SERVICOM as an institution and service delivery improvement as a goal.

Further evidence in favor of the complementarity of opportunities and energy is given by instances where a lack of opportunities for discussing performance partially undermined the channeling of reform energy into actual change and vice versa. This is perhaps seen most clearly in the Batho Pele reform, the symbolic importance and awareness of which was not matched by processes to enable civil servants to actually translate changed attitudes into changed practices. As one interviewee reflected:

> The problem is that it was, it was very politically driven, and oddly enough, it wasn't managerialist enough, in that it assumed that making government organizations more customer-oriented and providing better quality services to the public was just a matter of attitude of the public servants and they completely ignored all the other stuff which needs to be in place for an organization to provide better quality services. . . . The [assumption was that] service delivery here will improve if you have the right attitude. That was the problem with Batho Pele, it stopped there. It didn't make sure that all the systems and much more managerialist stuff was in place to enable public servants to provide better services.[21]

The cases where reforms provided processes but little energy are even more numerous, leading to committees, reviews, and reporting mechanisms that achieved perfunctory compliance at best.

Despite the importance of bringing together energy and opportunities, reformers faced a number of challenges and limitations in doing so that were common across various types of reforms and contexts. Even in the best cases, implementation and impact tended to be uneven across the civil service, as some organizations enthusiastically took up the opportunities while others treated them as impositions to be avoided or perfunctorily complied with. It was also difficult to sustain. High-level leadership and the energy that came from it could easily vanish overnight with changes in governments, political crises, or shifting priorities. Opportunities for discussion were often in the form of one-off workshops, evaluations, or support, which could sometimes yield one-off improvements in specific areas but were harder to channel into sustained improvement.

One way that reformers attempted to address the challenges of sustainability and uneven engagement was by trying to link these improvement opportunities to other processes that created pressure, accountability, and/or incentives for performance. Kenya's combination of Rapid Results Initiatives and Performance Contracting was a somewhat successful example of this; Ghana's efforts under CSPIP to combine organizational performance reviews with performance contracts for chief directors and Zambia's similar efforts under PSCAP were largely unsuccessful. However, the more processes that created opportunities for discussion and improving performance were linked (or perceived to be linked) to potential rewards or sanctions, the more resistance they met and the less genuine engagement they received. For example, South Africa's Management Performance Assessment Tool experienced gaming as well as pushback from politicians and bureaucrats when they tried to link scores to incentives or published results publicly, which eventually contributed to the system's demise. Similarly, the only instances in which individual-level leadership performance contracting systems were sustained over a period of multiple years were those where the incentives associated with measured performance were relatively diffuse (e.g., Kenya's Performance Contracting system had only a group bonus for members of the top-performing ministry but no formal punishments or individual incentives) or weak (e.g., Ghana's Chief Director Performance Agreement system in the mid- to late-2010s, which had small and mainly symbolic awards for top performing chief directors). So institutionalizing top-down energy through high-powered formal incentive and accountability systems was rarely a successful route to sustained performance improvement.

The provision of financial resources to organizations was sometimes used to try to elicit energy from the organization. But this proved to be a double-edged sword: Resources generated interest from organizations in reform efforts, but this interest often became oriented around the resources themselves rather than on how they could be used to improve performance. This effect was compounded by accountability and budget-disbursement pressures that were more severe for donor-funded projects (but not solely restricted to them). For example, the Performance Improvement Funds (PIFs) in Ghana under CSPIP and Zambia under PSCAP each succeeded in not only generating some innovative and impactful initiatives but also led to many proposals that were only loosely linked to performance improvement and often focused purely on purchasing assets like cars or computers. In contrast, Kenya's Rapid Results Initiatives—which aimed at sparking similarly specific, short-term, innovative changes as Ghana and Zambia's PIFs—typically did not come with additional financial resources and largely avoided these distortions. Like carrots and sticks, resources were as often a hindrance as a help to efforts to combine and sustain energy and opportunities.

There was no simple formula for predicting when, where, and how energy and opportunities for discussing and making performance improvements successfully came together. Reforms that had formally similar designs and contexts often led to different outcomes in terms of whether these mechanisms of success emerged and amplified one another. Similarly, there were no design features of reforms or characteristics of reform context that guaranteed that these mechanisms would emerge or combine. Energy and/or opportunities could emerge in various ways and in different contexts as well as fail to do so.

On the one hand, this ambiguity is frustrating for reformers and academics alike, as it reinforces the idea that there are no simple, replicable, easily transferable, magic-bullet answers to the question of how to design and implement systemic performance-oriented reforms. On the other hand, it also seems appropriate and obvious that there wouldn't be a simple answer to the challenge of getting tens or hundreds of thousands of civil servants spread across hundreds of organizations to simultaneously undertake tasks that can't always be specified in advance or measured after the fact. Complex challenges like this don't usually have easy solutions.

However, the fact that there aren't simple, guaranteed-to-work solutions doesn't mean that there aren't patterns and empirical regularities to learn from. This chapter makes clear that when reforms did succeed in changing unformalizable behaviors, it was generally through one or of the two mechanisms of creating energy and creating opportunities—ideally both. There were also regularities in the proximate factors that enabled each mechanism to operate: Energy was successfully created and spread when civil servants' individual and collective expectations about the possibility and likelihood of meaningful change were shifted, and opportunities were successfully created when civil servants were given space and prompting to communicate with one another about goals, roles, feedback, and ideas for improvement. Neither mechanism could be forced or imposed top-down on a civil service, but both appeared in many diverse types of reforms and contexts. Together, they comprised the two main channels through which reforms managed to positively change bureaucratic behaviors and improve performance.

• • •

The end of this chapter begins our transition from part II of the book, on the diagnosis of how past reforms failed and succeeded, to part III of the book, on the prescription for how reformers can maximize the likelihood that their reforms are successful. It has argued that there were two main success mechanisms through which reforms consistently had positive impacts on behavior and performance: creating opportunities for civil servants to discuss performance and how to improve it, and creating energy for change. These mechanisms could each operate independently but reinforced each other when they were both present. While these mechanisms faced a range of limitations and challenges, they nonetheless recurred across a wide range of reform types and contexts.

One theme that unites part II's analysis of the mechanisms of failure and the mechanisms of success is the complex relationship between formal and informal processes and actions. On the one hand, the overfocus on changing formal rules and structures as an objective of reform and a way to improve performance was a clear and consistent mechanism of failure, and the actions sparked by the mechanisms of success were decidedly nonverifiable. On the other hand, these informal actions were still often undergirded or prompted by formal processes—or constrained and squeezed out by them. So any attempt to think about how reformers should approach designing and implementing reforms needs to grapple with both types of behavior and practices.

A second theme uniting the chapters in this section was the importance of civil servants' expectations about how their colleagues and supervisors will react to a reform effort and, thus, of the credibility (or not) of reform efforts. Expectations and credibility mattered both for whether civil servants complied with formal processes and for whether they undertook important but unformalizable actions to give meaning to these formal processes. So the implementation and impact of a reform were determined not just by its content and its designers' intent but by how it was perceived by the civil servants whose behavior it aimed to influence. Common approaches to designing and implementing reforms could inadvertently undermine this credibility—as could external factors like leadership turnover, the uncertainty of future funding, or a lack of support from key stakeholders. So finding ways to maximize the credibility of reform efforts and change civil servants' expectations is also crucial for thinking about to learn from these past experiences and approach future reform efforts.

Once again, these patterns are not restricted to these six countries or, indeed, to Africa in general. For example, the public administration scholar Donald Moynihan concluded in his study of performance information use in the United States that while organizational performance management systems almost never functioned as intended—they produced too much information, performance was too ambiguous to interpret, and elected officials ultimately didn't care very much about the formal performance ratings—they still often had positive impacts within agencies by enabling "interactive dialogue" about objectives and performance.[22] In their review of worldwide evidence on staff appraisal procedures for the World Bank, Schnell and colleagues recommend "that managers separate conversations focused on employee development from formal performance conversations that have primarily accountability and control functions" because linking conversations to rewards and sanctions triggered defensiveness, exaggeration, conflict avoidance, and behavioral biases among both supervisors and supervisees.[23] I imagine that most people reading this book have their own experiences of how useful conversations about goals, roles, and performance can be, as well as how difficult it is to be open and honest in such conversations when rewards and sanctions are on the line.

One obvious potential takeaway from this part of the book is this: Reformers should seek to minimize the mechanisms of failure and maximize the mechanisms of success. To do less of what hasn't worked and more of what has. For example, annual staff appraisal systems consistently failed to serve as the incentive delivery systems that reform designers envisioned, but they sometimes created useful opportunities for supervisors and supervisees to discuss goals, roles, and performance feedback. But the more burdensome these systems were and the more they threatened or promised incentives attached to these performance evaluations, the more that they were gamed or treated as tick-box exercises. What, then, would it look like to design annual appraisal systems that abandoned the failed premise of linking carrots and sticks to measured individual performance and instead sought to maximize energy and opportunities for supervisors and supervisees to share information, come up with ideas to improve performance, and put them into action? While the operation of these mechanisms and the interaction between them is, of course, more complex than just saying "less of A, more of B," these patterns of failure and success nonetheless provide a valuable starting point for thinking about how to design and implement more impactful reforms.

These are the questions that the next part of the book turns to. Building on the reform histories in the appendix and the analysis of past reforms from this part of the book, part III begins by building a theory of reform as an *ongoing process of catalyzing continuous improvement in actual practices*. It explores what this idea might mean not only in the abstract but also grounds it in past reforms that have exemplified aspects of this approach, both in Africa and around the world. While success can never be guaranteed for complex, systemic, large-scale reforms in an uncertain world, for a range of performance improvement objectives, this approach might offer reformers a better chance at success than past efforts have had. That's chapter 7. Chapter 8 then takes an in-depth, practice-oriented look at one version of what this theory of reform as process might look like. To do so, it examines how Nana Agyekum-Dwamena took a dramatically different approach to reform after he took office in 2014, an approach that exemplified this alternative approach to reform in many ways. Finally, the book closes by reflecting on the challenges and potential limitations of this approach and of the book's analysis, as well as how the optimal approach to reform might vary depending on reformers' objectives and contexts. With that in mind, let's move on to this third and final part of the book.