WORKING PAPER

Delivery Approaches to Improving Policy Implementation: A Conceptual Framework

April 14, 2021

Martin J. Williams
Clare Leaver
Zahra Mansoor
Dana Qarout
Anna Bilous

Blavatnik School of Government
University of Oxford

Karen Mundy
Minahil Asim
Sheena Bell

University of Toronto

Acknowledgements

Please cite this paper as:

Williams, Martin J., Clare Leaver, Karen Mundy, Zahra Mansoor, Dana Qarout, Minahil Asim, Sheena Bell, and Anna Bilous. (2021). "Delivery Approaches to Improving Policy Implementation: A Conceptual Framework." DeliverEd Initiative Working Paper. Education Commission and Blavatnik School of Government.

About DeliverEd

The DeliverEd Initiative was launched in 2019 and aims to strengthen the evidence base for how governments can achieve their policy priorities through delivery approaches. Globally, more than 80 countries have used delivery approaches to achieve better outcomes with reform and policy implementation. But to date, little empirical evidence, especially from low- and middle-income countries, exists on the effectiveness of delivery approaches and what design choices and contextual features contribute to this impact. The DeliverEd Initiative aims to fill this evidence gap and create a better understanding of what practices leaders can adopt to improve their policy delivery and reform efforts.

The Education Commission is leading this initiative in partnership with the Blavatnik School of Government, the University of Toronto, and with funding from the UK Foreign, Commonwealth and Development Office. Other partners include IDEAS, Georgetown University's gui2de, and the World Bank.

For more information about DeliverEd and to view other related research products, please visit www.educationcommission.org/delivered-initiative

Table of Contents

1. Introduction

Governments all around the world are adopting ambitious goals and reforms aimed at improving the quality-of-service delivery, particularly in the education sector. However, they often struggle to translate these reform intentions into tangible action because these high-level intentions must be carefully formulated into an actionable plan, transmitted through a complex and multi-sited bureaucratic system, and then put into practice by frontline bureaucrats. Achieving rapid changes in the functioning of service delivery bureaucracies can be challenging due to the potential bureaucratic inertia and resistance, as well as the complexity, coordination, discretion, and innovation required to achieve systemic change. Understanding how to improve bureaucratic functioning and policy delivery has thus emerged as one of the main challenges facing governments worldwide.

In the past 20 years, delivery units, delivery labs, and other similar approaches have been adopted by some governments aiming to implement reforms or achieve high-level targets. These *delivery approaches* have been adapted and adopted in various forms by dozens of governments worldwide, with their use in the education sector being particularly widespread. The primary objective of these delivery approaches is to improve policy delivery by changing bureaucratic functioning. Research under the DeliverEd project seeks to provide a new body of rigorous evidence on the potential effectiveness of delivery approaches for achieving these goals, led by an ongoing set of in-depth country-level case studies of the particular delivery approaches in Tanzania, Pakistan, Ghana, Jordan, and India.

The purpose of this paper is to outline a preliminary conceptual framework and set of definitions that can be used by each country's research teams to establish a common vocabulary and set of research questions to harmonize their independent analyses. In particular, this paper focuses on establishing a definition and scope for what constitutes a delivery approach, identifying a set of design features that are common to many delivery approaches, and highlighting a selected number of key dimensions along which the design of delivery approaches can differ.

As with any management tool, these different design choices sometimes entail trade-offs between different goals and activities, which in turn also influence their fit with different contexts. Understanding these different approaches and their associated trade-offs is an important step to understanding how, when, where, and why delivery approaches might be most usefully deployed, as well as for relating delivery approaches to other ways in which governments can try to improve service delivery. By design, this framework highlights only a selected few of the many potential determinants of delivery approaches' effectiveness to strike a balance between providing a common set of terms to promote consistency among country-level

empirical studies while also leaving flexibility for each country study to explore the factors that emerge as most important in their contexts. This conceptual framework is thus intended to evolve over the course of the research in response to evidence emerging from the country case studies.

For the purpose of this project, we define a delivery approach as *an institutionalized unit or structured process within a government bureaucracy that aims to rapidly improve bureaucratic functioning and policy delivery by combining a set of managerial functions in a novel way to shift attention from inputs and processes to outputs and outcomes.* Although delivery units are well known, there are many ways to design and operationalize them, and delivery units per se are just one species of a broader class of delivery approaches that governments can adopt, as many of the functions performed by delivery units can (and often are) combined in different ways and can be carried out by different structures within a system – including through standard bureaucratic structures. If a minister asks the question, "How should I improve policy delivery in my sector?', then adopting a delivery approach might be one answer. But what is the range of such approaches from which the minister might choose, and which of these approaches might be best suited for different purposes and contexts?

Section 2 briefly surveys conceptual approaches to and existing evidence on delivery approaches. We find that most existing conceptual frameworks are normatively focused (i.e., what *should* delivery approaches do), which limits their suitability for use as a basis for analysis of their effectiveness, and that there is little empirical evidence on their effectiveness that meets academic standards of rigor. In Section 3, we build on this existing literature to develop our own definition of a delivery approach and what it seeks to achieve that is broad enough to encapsulate the wide range of forms delivery approaches take, while still having boundaries that limit the scope of the concept.

Section 4 then embeds this understanding of delivery approaches within a broader system-level theory of change for how delivery approaches fit int efforts to improve the performance of the education sector as a whole. This stylized theory of change includes the goals of and inputs to a delivery approach, which are 'upstream' or 'prior to' the operation of a delivery approach; the functions undertaken by the delivery approaches itself; the downstream changes in bureaucratic functioning and policy delivery that delivery approaches seek to directly affect; the final outcomes such as improved teaching quality and student learning that are the ultimate goals of delivery approaches, but which delivery approaches can only affect through improved bureaucratic functioning and policy delivery; and features of political and institutional context that can moderate the effectiveness of different delivery approaches. Rather than being oriented at suggesting how best to design or operate a delivery approach,

in the context of this preliminary conceptual framework, the purpose of this theory of change is to help establish a common understanding of how these selected features and effects of delivery approaches relate to each other. We illustrate some potential connections and feedback loops among these categories and note numerous important factors that are not highlighted in our stylized theory of change, but which may be important for country-level research teams to investigate.

Section 5 discusses the range of functions undertaken by delivery approaches in more detail, classifying them into five main categories: prioritization and target setting, monitoring, and measurement, leveraging political sponsorship, accountability and incentives, and problem-solving. We also discuss other potential functions that do not fit into these five categories, as well as non-design elements like leadership and culture that may also influence a delivery approach's effectiveness. Within this, we identify two key pathways to impact which most approaches combine to various extents. What we term 'Pathway A' is focused on leveraging tools of accountability and incentives (financial or non-financial) to increase effort among bureaucrats and school management. In contrast, 'Pathway B' relies more on efforts to stimulate organizational learning, coordination, and problem-solving among bureaucrats and school management. This distinction builds on a longstanding body of theory and debates in public administration on management and motivation (e.g., Friedrich, 1940; Finer, 1941; Simon, 1983; Rose-Ackerman 1986; Wilson 1989; Rainey & Steinbauer, 1999; Carpenter, 2001; Gruening, 2001; Le Grand, 2003; Chun & Rainey, 2005; Duflo et al., 2012; Andersen & Moynihan, 2016; Andrews et al., 2017; Honig, 2018; Rasul & Rogger, 2018; Rasul et al., 2020) and management studies (e.g., McGregor's classic [1960] distinction between 'Theory X' and 'Theory Y') as well as discussions of key impact pathways in the literature on delivery approaches in particular (e.g., Barber, 2015; Nordstrum et al., 2017; O'Malley, 2019). We describe these two stylized pathways in more detail and relate them to observed empirical cases. In practice, most actual delivery approaches combine both pathways to some degree, and there are several common functions across both approaches. However, the conceptual distinction is useful because these two pathways represent different understandings and hypotheses about the underlying causes of poor policy implementation and because some delivery approaches rely more heavily on one pathway than another. As discussed in Section 5, these pathways may also interact with each other – either in ways that complement or potentially even undermine each other. A core purpose of this conceptual framework is to make these alternative designs, associated theories of change, and links to goals and context more explicit, to facilitate empirical study by the DeliverEd project's country-level research and build evidence on whether and why some styles of delivery approaches might be more effective in certain contexts or for certain types of goals.

Section 6 then discusses some key conceptual distinctions about the upstream, downstream, and contextual factors that form the remainder of the theory of change

for delivery approaches. Section 7 discusses how a range of research questions can be couched within this theory of change. Having a unified conceptual framework is crucial to be able to formulate hypotheses and interpret the disparate findings of empirical studies that focus on different cases and may find different results. However, Section 7 stops short of formulating hypotheses that are precise enough to be rigorously and empirically analyzed because this would require adopting specific theoretical stances and relating them to specific empirical cases – which is beyond the scope of this framework paper but will be addressed in a future methodological paper. Section 7 also highlights a range of additional variables that are not highlighted in the limited conceptual framework, but which may be fruitful for country-level research teams to investigate, including a wider set of contextual characteristics, aspects of the design and operation of delivery approaches, outcomes, and links to other stakeholders. Section 8 concludes the paper by summarizing the framework as well as our plans for its implementation.

While our conceptual framework applies to delivery approaches at all levels and sectors, we focus our discussion on the use of delivery approaches in the education sector. This helps us to establish tangible examples, without diluting the generality of the framework. Whether and how delivery approaches differ between the education sector and other sectors is an interesting empirical question to which this conceptual framework can be applied.

2. Existing Theory and Evidence on Delivery Approaches

2.1 Background and Prior Conceptualizations

The most widely known type of delivery approach is the *delivery unit.* The first models of the delivery unit emerged in the 1990s and early 2000s, with the New York City CompStat policing program and the United Kingdom's Prime Minister's Delivery Unit (PMDU) widely cited as early examples (Gold, 2017). These specialist units aimed to focus political and bureaucratic attention on eliminating obstacles to efficient public service delivery through defining clear targets and instituting data collection mechanisms that support improved performance management routines and problem-solving (Barber et al., 2011; Kohli et al., 2016; Delivery Associates, 2018). Following these early examples, there has been a proliferation of delivery units around the world, particularly in the education sector, with a range of variations in their scope (e.g., focus on a few versus many priorities), form (e.g., nature of staffing structures and institutional set-up), and the types of managerial approaches (Alessandro et al., 2014; Andrews, 2014; Todd, 2014; Gold, 2014; Harrison, 2016; Gold, 2017; Lafuente & Gonzalez, 2018). Numerous existing studies have documented the adoption, design, operation, and successes and challenges of delivery units, as well as laid out recommendations for policymakers. This includes several texts written by key actors involved in the operation of some of these units (e.g., Barber et al., 2011; Andrews, 2014; Barber, 2015; Kohli et al., 2016), as well as a range of academic and practice-oriented case studies of particular units, dynamics, and episodes (e.g., Hood & Dixon,

2010; Scharff, 2012; Scharff, 2013; Simson, 2013; Gold, 2014; Freeguard & Gold, 2015; Harrison, 2016; Hart, 2017; Gold, 2017). While a full review of the key insights and debates emerging from this literature is beyond the scope of this working paper, some of the key lessons that emerge include: the importance of establishing routines for measuring and discussing performance (e.g., Todd, 2014; Barber et al., 2011; Andrews, 2014; Barber, 2015; World Bank, 2017; Hart, 2017; Delivery Associates, 2018; Lafuente & Gonzalez, 2018); the challenges of and various approaches to connecting parts of government that rarely communicate (e.g., Scism, 2015; Lafuente & Gonzalez, 2018); the potential interactions between the accountability-focused and problem-solving, focused pathways to improving service delivery (e.g., Todd, 2014; Nordstrum et al., 2017; O'Malley, 2019); and the importance of leadership, appropriate staffing, and organizational culture within the delivery approach itself, as well as managing relationships with both political sponsors and downstream service delivery actors (e.g., Shostak et al., 2014; Andrews, 2014; Barber, 2015; Gold, 2017; Lafuente & Gonzalez, 2018; O'Malley, 2019). This rich body of literature has informed the questions and hypotheses of the DeliverEd project as a whole and serves as a strong foundation both for this paper's conceptual framework and for the ongoing country-level research.

Our study of delivery approaches is also informed by a vast range of literature in public administration, education, economics, management, and political science on bureaucratic performance and service delivery. Selected highlights from this literature include debates on failures in policy implementation and frameworks that characterize them (e.g., Pressman & Wildavsky, 1979; World Bank, 2004; Pritchett, 2015); specific service delivery failures that motivate the use of delivery approaches (e.g., Tiernan, 2006; Lindquist, 2006; Lindquist & Wanna, 2010; Hood & Dixon; 2010; Chambers et al., 2012); the distinction, and potential trade-offs, between managerial approaches that relate to accountability and/or incentives versus collaborative problem-solving and organizational learning for bureaucratic management (e.g., Simon, 1983; Rose-Ackerman, 1986; Wilson 1989; Rainey & Steinbauer, 1999; Carpenter, 2001; Gruening, 2001; Chun & Rainey, 2005; Duflo et al., 2012; Andersen & Moynihan 2016; Andrews et al., 2017; Rasul & Rogger, 2018; Rasul et al., 2020); and discussions on the role of leadership, management practices, and organizational culture in shaping bureaucratic functioning (e.g., Grindle, 1997; Bloom et al., 2007; Bloom et al., 2012; Gibbons & Henderson, 2013; Bloom et al., 2014; McDonnell, 2017, 2020). Several authors have previously attempted to classify the key elements, principles, or functions of delivery approaches. Table 1 presents a selection of prominent classifications; it indicates there is no universally agreed-on definition of delivery units (or of the broader concept of delivery approaches). Most existing classifications tend to be aimed more at making recommendations about the elements of successful delivery approaches. This contrasts with the purpose of this paper's conceptual framework, which is to identify potential differences in the design and operation of different delivery approaches to facilitate empirical study. Some definitions of delivery approaches, such as Gold's (2014) and Shostak et al.'s (2014), describe delivery units as one tool used

by the center of government to achieve a limited number of outcomes that are a top priority for the government rather than a delivery approach for the system as a whole. Thus, in some cases, delivery units may exist as a feature of a broader delivery approach, with the unit focused specifically on carrying out a subset of delivery functions such as data analysis, problem-solving, and escalation of implementation bottlenecks, while in other cases the unit itself may be the entirety of the delivery approach adopted by a government. In Sections 3 and 4 below, we take this latter approach to creating a definition and typology for understanding delivery approaches.

Table 1: Selected Typologies of Delivery Unit Functions

Barber et al. (2011)	World Bank (2010a, 2010b)	PEMANDU Associates (n.d.)	Kohli & Moody (2016)	Todd & Waistell (forthcoming)	Gold (2014)	Shostak et al. (2014)
1. Setting direction and context. 2. Establishing clear accountabilities and metrics. 3. Creating realistic budgets, plans, and targets. 4. Tracking performance effectively. 5. Holding robust performance dialogues. 6. Ensuring actions, rewards, and consequences.	1. Political signaling around key priority areas. 2. Tracking progress through a monitoring and reporting system. 3. Political sponsorship of reform areas. 4. Establishing forums for problem-solving, coordination, collaboration, and innovation. 5. Ensuring that Ministers and senior staff know they are held accountable for the delivery and will face consequences accordingly.	1. Setting strategic visions and targets. 2. Developing implementation plans in consultation with all stakeholders involved in the delivery chain. 3. Soliciting feedback from clients (citizens/ beneficiaries) and other stakeholders to revise implementation plans. 4. Communicating progress regularly and widely to signal transparency and accountability for results (including to the public). 5. Setting targets and review them periodically based on analysis of the data. 6. Establishing monitoring and problem-solving routines. 7. Validating and verifying progress through external audits.	1. Developing delivery plans with clear task owners and data collection processes, targets, and benchmarks. 2. Analyzing data to produce relevant and important insights to improve delivery. 3. Partnering with other units and departments involved in the delivery chain, especially frontline practitioners. 4. Cultivating a culture in government geared towards rigor in problem-solving, mutual trust between leaders and their team, and transparent communication about progress and results. 5. Maintaining institutional independence from ministries and other public agencies to uphold objectivity.	1. Prioritizing targets and ensuring sufficient resource allocation for implementing these priorities. 2. Setting up data collection, analysis, and information sharing routines to inform decisions. 3. Analyzing delivery challenges by engaging various stakeholders (to understand incentives and perception of agents). 4. Maintaining open communication channels with delivery agents and public and ensuring accountability relationships with responsible stakeholders. 5. Securing and signaling senior leadership support and commitment to priorities through communication and routine monitoring and reporting.	1. Planning implementation and assessing the feasibility of proposed plans. 2. Monitoring progress towards priorities through regular follow-up routines and field visits, typically carried out by "delivery units." 3. Managing projects through the deployment of talented staff and facilitation of cross-departmental coordination at the center of government. 4. Setting up and utilizing evaluation routines that generate evidence at various stages of the implementation process that feed into planning and decision-making about implementation trajectories and priority targets.	1. Improving outcomes for citizens by focusing on a limited number of priorities for implementation. 2. Resolving barriers to and disruptions of implementation/ service delivery based on monitoring data. 3. Understanding and strengthening the systems' and agents' ability to improve implementation processes and skills.

2.2 Empirical Studies

There exists a plethora of policy literature documenting claimed successes and challenges of delivery units worldwide. However, much of this literature is based on before-vs.-after comparisons of indicators or interviews with involved actors. There is little evidence that evaluates the impacts of delivery approaches rigorously, either by establishing a counterfactual or by conducting the type of qualitative process-tracing and triangulation that can interrogate these impact claims more deeply. Nevertheless, this existing body of evidence provides some empirical support for the potential effectiveness of delivery approaches, as well as the range of political and bureaucratic dynamics in which they are embedded.

Many delivery units have been associated with significant improvements in indicators associated with priority targets. For example, following the establishment of the Performance Management and Delivery Unit (PEMANDU) in Malaysia, street crime dropped by 35%, the percentage of Malaysians reporting high levels of satisfaction with the government's efforts in reducing corruption increased by 20 percentage points, and an additional 2 million Malaysian citizens benefited from improved public services such as water sanitation, electricity, roads, and housing (CPI, 2016b). Chile witnessed a drop in crime victimization from 34% in 2009 to 25% in 2013 (Lafuente & Gonzalez, 2018). In Colombia, theft of cell phones declined by 12% in a single year (Lafuente & Gonzalez, 2018). Between 2011 and 2015, the state of Pernambuco in Brazil experienced a boost in its ranking on the standardized high school achievement test, going from 16th place out of 27 states to first place; this Brazilian state also ranked first in lowest high school dropout rates, an improvement from being ranked the 11th out of 27 (Lafuente & Gonzalez, 2018). Similarly, hospital waiting times in the UK reduced significantly for 10,000 patients who no longer had to wait a year or more for surgeries (CPI, 2016a). The state of Maryland in the United States experienced a 14% reduction in murder rates, and Baltimore's local government secured a much-needed quick-win for the mayor by leveraging their online data analytics dashboard, CitiStat, to expedite solving cases by law enforcement agencies by clearing an accumulation of over 24,000 DNA samples in need of lab analysis (Freeguard & Gold, 2015). Other improvements have been documented in the areas of childhood immunization, street crime, public infrastructure, school ranking, infant mortality, and hospital waiting times (Gold, 2017; Lafuente & Gonzalez, 2018; Alessandro et al., 2014, Baars et al., 2014; Alari & Thomas 2016; CPI, 2016a, CPI, 2016b). In addition to their impacts on final outcomes like these, other studies have documented delivery approaches having effects on intermediate outcomes like improved coordination and performance orientation (Scharff, 2012; Andrews et al., 2017). Some studies have pointed out potential evolutions in the character of delivery approaches over time. For example, Scism (2015) discusses the expansion of functions beyond monitoring and performance management towards a broader range of forms of engagement with stakeholders, communication, and seeking of policy input. Similarly, various authors

have noted the gradual spread of the institutional location of delivery approaches from the center of government towards various decentralized delivery management structures at the sectoral, state, or provincial level in countries like Tanzania, South Africa, the United States, Canada, Brazil, and the UK (Gold, 2017; Langford and Roy, 2008; Alessandro et al., 2014). Other studies have claimed a link between the presence of a delivery approach and the uptake of innovative practice in government (Gold, 2017; World Bank; 2017; Freeguard & Gold, 2015).

Another empirical pattern is the often-short-lived nature of delivery units which Gold (2017) observes often correspond to the leadership term of a particular president, prime minister, or minister. This is perhaps a consequence of most delivery approaches' heavy reliance on political sponsorship from high-level political figures. However, Delivery Associates (2018) argue that short lifespans are not necessarily indicative of success or failure; rather, the lifetime of a delivery unit should be related to the timeline set for desired outcomes. The extent to which short lifespans are caused by leadership turnover as opposed to intentionally time-bound targets has not been investigated empirically. Another trend observed in delivery approaches around the world is the role of international development and donor agencies. International development and donor organizations are increasingly invested in promoting such delivery approaches, especially in low- and middle-income countries (Gold, 2017).[1]

The policy-oriented and limited academic literature that exists on delivery approaches also highlights some common challenges. These challenges include: a) dealing with political transitions (e.g., Scharff, 2013; NCHRD, 2016); b) limited influence over delivery chains (Freeguard & Gold 2015; Delivery Associates, 2018); c) entrenched operating cultures and practices (e.g., Barber et al., 2011; Scharff, 2012; Barber, 2015; Gold 2017); d) measuring relevant outcomes (Barber et al., 2011; Aviv, 2014; Andrews, 2014; Barber, 2015; Freeguard & Gold 2015); and e) citizen engagement to maintain political support and solicit inputs (e.g., Shostak et al., 2014; Andrews, 2014; Barber, 2015; Freeguard & Gold, 2015; Delivery Associates, 2018; CPI, 2016a). While the conceptual framework outlined in this paper does not speak directly to these issues, it nonetheless provides a foundation from which they can be explored in greater depth in country-level empirical research.

3. Definition and Scope of Delivery Approaches

For the purposes of the DeliverEd project, we define a delivery approach as *an institutionalized unit or structured process within a government bureaucracy that aims*

---

[1] According to Gold (2017), the most prevalent organizations include UK's Department for International Development (DFID), the Swedish International Development Cooperation Agency (Sida), World Bank, UNDP, Inter-American Development Bank (IDB), and United States Agency for International Development (USAID).

to rapidly improve bureaucratic functioning and policy delivery by combining a set of managerial functions in a novel way to shift attention from inputs and processes to outputs and outcomes. This definition includes not just delivery units *per se* but also a range of other institutional forms such as delivery labs, innovation units, and reform teams (although not all institutions with these terms in their name fall within our definition). This definition is admittedly and deliberately broad, so some further clarification is necessary to distinguish delivery approaches from other types of reforms, structures, and interventions. These distinctions also serve a substantive purpose, as they help to clarify what delivery approaches do and what alternative approaches for achieving similar objectives exist.

One conceptual challenge is that the management tools and functions utilized by delivery approaches are not exclusive to delivery approaches. Indeed, these are core functions of all bureaucracies, and delivery units and labs often have remarkably similar goals and use similar tools as those of existing structures like budget offices, management boards, monitoring and evaluation units, or prime minister's offices (Lindquist, 2006). What is the relationship of delivery approaches to these widespread structures? Or even more generally, what is the relationship of delivery units to other institutional efforts to improve service delivery (e.g., decentralization), or even to non-governmental accountability efforts like community monitoring of schools? In the remainder of this section, we use our definition of delivery approaches (*'an institutionalized unit or structured process within a government bureaucracy that aims to rapidly improve bureaucratic functioning and policy delivery by combining a set of managerial functions in a novel way to shift attention from inputs and processes to outputs and outcomes'*) to make a series of distinctions about the boundaries of the delivery approach concept and its relationship to other institutional forms and interventions.

First, delivery approaches may not always introduce new practices and processes, but instead, bundle a set of pre-existing functions together in a *novel configuration*. In some cases, a delivery approach may genuinely introduce new practices (e.g., high-stakes targets, new data collection, more frequent stakeholder convening), but in other cases, a delivery approach may seek to achieve its effect simply by *combining* existing practices from throughout the bureaucracy into one unit. In this case, the hypothesis would be that bundling these functions into one part of the bureaucracy would have a greater effect or improve their execution, compared to spreading these functions across specialized parts of the bureaucracy (e.g., monitoring by monitoring units, accountability from ministerial offices, targets from planning directorate). While delivery units typically involve centralization of these functions into a single institution close to political leadership, more distributed or decentralized delivery approaches that utilize existing bureaucratic structures or exist outside the core executive are also possible; we further discuss the potential for variation in institutional configurations in Section 4.

Second, we identify delivery approaches as those that are *spearheaded by government*. Initiatives that are exclusively externally led by non-government organizations (NGOs), donors, or local communities are not best understood as delivery approaches.[2] Relatedly, delivery approaches are primarily *managerial* approaches, utilizing tools that are available to executive leaders and bureaucracies, rather than community-led approaches to service delivery. While this does not preclude collaborative processes in how delivery approaches are implemented, delivery approaches are fundamentally government-led and utilize managerial tools available to political or bureaucratic leadership. For this same reason, interventions that seek to improve service delivery by transferring the locus of responsibility or resources (e.g., decentralization, privatization or contracting out, provision by NGOs, direct transfers to households, community participation) are not best understood as delivery approaches (even if they might have similar goals of improving service delivery outcomes).

Finally, we focus on approaches that are institutionalized and aim to transform implementation and performance in a broad-based and lasting fashion, and which focus on improvement in achieving outputs and outcomes (as opposed to on spending resources or undertaking processes). This excludes initiatives such as one-off training interventions, cash transfers, narrow time-bound projects, or performance-based incentive programs. It would also exclude most project management units (of the form that are often associated with donor-funded projects) for which the primary objective is to deliver resources or execute a specific, pre-specified set of activities rather than to catalyze a broader transformation in public service delivery since these are primarily input- and activity-focused. It would also exclude initiatives like performance audit units or better budget management which are process-focused rather than output- or outcome-focused. However, this would include delivery approaches that are created to be temporary in nature, for example, for use in responding to a specific crisis or enacting a broad program of bureaucratic reforms.

## 4. A Theory of Change of Delivery Approaches

Having defined delivery approaches, the next question is: What role do delivery approaches (DAs) play in the broader system of education service delivery? Put another way, how does the creation of a delivery approach fit within the broader delivery chain involved in providing services to the population? A clear understanding of this is necessary to both clarify what other variables influence DAs' design and operation and are influenced by them, and to formulate hypotheses about these effects for empirical testing. Our conceptual framework for this is positive rather than normative, in the sense that it simply attempts to describe and relate the relevant

---

[2] This does not exclude delivery units that are co-managed by consultants or contracted out to consultants. This is because in both these instances, the government remains the client.

factors to each other, rather than making any recommendations about what role DAs *should* play or how they should be designed. It is also a highly stylized framework in the sense that it focuses on establishing relationships among a minimal set of variables for simplicity and clarity and, to preserve flexibility for country research teams, to adapt it to the particularities of their own context.

Figure 1 depicts our high-level theory of change for DAs. We focus on the case of the education sector and DAs established in the center of government, but the framework can be generalized and adapted to any service delivery sector or to DAs that operate at different levels of government. Figure 1 distinguishes among five key sets of variables: 1) functions undertaken by the DA (blue), which represent what the DA itself actually does to try to achieve its impacts and which we distinguish into two pathways that bundle functions together, described in greater detail below; 2) goals and inputs of the DA (orange), which are decisions taken by higher-level authorities prior to the operation of the DA during its design and set-up; 3) changes in bureaucratic functioning and policy delivery at the national, sub-national, and school administration levels (green), resulting directly from the DA undertaking its functions; 4) changes in the final outcomes of interest (yellow) which are themselves the results of the better bureaucratic functioning that might result from the DA's operations, but which are not themselves directly affected by the DA; and finally, 5) these decisions, actions, and outcomes all occur in a political and institutional context (gray) which can directly affect each set of variables, as well as interact with them in determining their consequences for downstream variables. While the overall theory underlying DAs is that the introduction of a DA into this delivery chain leads to a change in final outcomes related to frontline service delivery (in the education sector case, improved teaching and better student learning), understanding how and whether this is achieved requires consideration of each set of variables.

Figure 1: A Theory of Change of Delivery Approaches
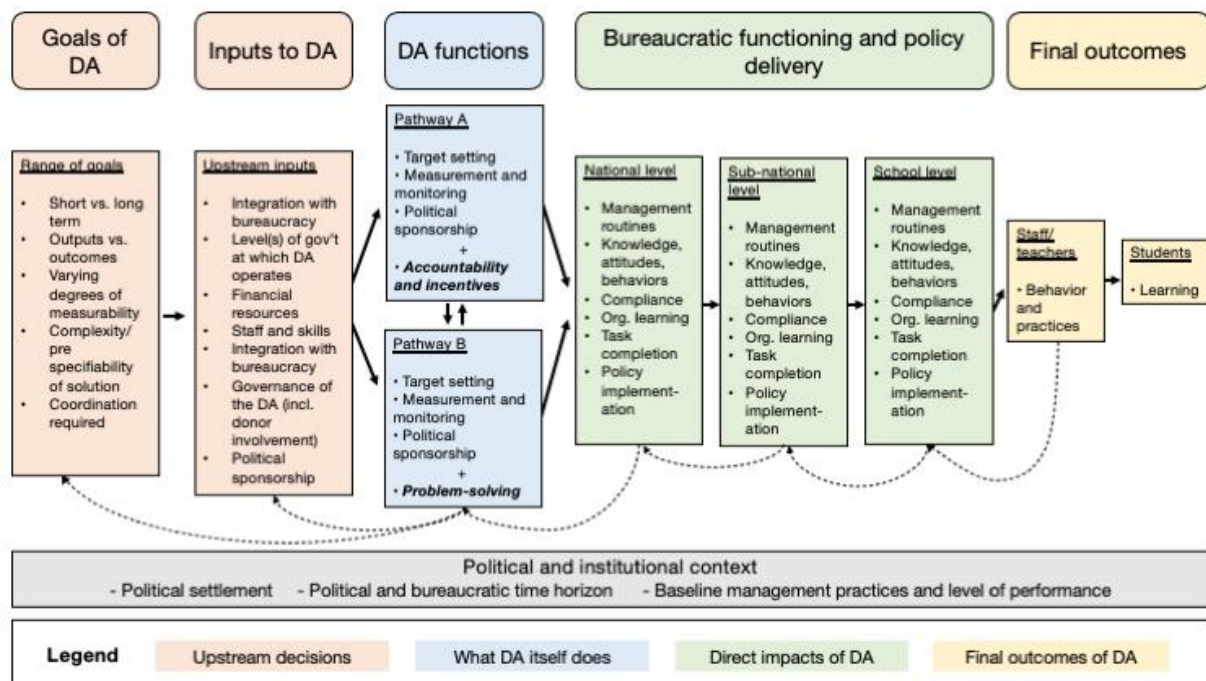
As well as making explicit the overall theory of change that underpins the adoption of DAs, this figure also illustrates a set of ways in which the adoption of a DA might *fail* to lead to improved learning outcomes. Once a policy target or reform is adopted by a government and a DA is introduced to improve its implementation, there are five potential gaps that could result in no improvements in learning outcomes. First, the DA may not receive the required inputs to undertake its functions. Second, the DA may have the necessary inputs but carry out its functions poorly, for example due to poor leadership, culture, or internal management. Third, the DA may execute its functions as intended, but its implicit hypothesis about which functions will trigger improvements in bureaucratic functioning and implementation at lower levels of government might be wrong, leading to no change or even deterioration in performance, if for example the functions are ineffective or incoherent. Fourth, the DA might successfully catalyze improvements in bureaucratic functioning and policy delivery, but the causal link between the policy and its impact on teacher practices may not exist as hypothesized. This could also be conceptualized as 'choosing the wrong policy' at the goal stage. Finally, the policy may get implemented and change teacher practices as intended, but these teacher practices may not causally lead to improved learning outcomes. Recalling our definitions from Section 2, we would term the second and third types of failures as 'implementation failures', which the DA was intended to solve.

This figure is, by necessity, a vast simplification of a complex system. We make these simplifications to focus attention on the key categories and linkages, not to deny the existence of others. We illustrate the DA functions in a linear framework that

corresponds to standard bureaucratic hierarchies. However, in reality, understanding causal effects and implementation failures often require simultaneous analysis across different levels. There are also potential feedback loops embedded into the system either by design or by accident. Figure 1 illustrates some particularly important ones, but there are many other types of feedback that might be possible. In addition, we focus our attention on the bureaucratic aspects of service delivery, setting aside the interplay between a) politicians and citizens and b) citizens and frontline service deliverers, which are also important determinants of service delivery outcomes and are highlighted in other conceptual frameworks (e.g., World Bank, 2004; Pritchett, 2015). Where these interact with delivery approaches, they can of course be brought back into the analysis (such as when politicians use delivery approaches to try to leverage citizen pressure onto frontline agents); but Figure 1 focuses more narrowly on the main mechanism through which most delivery approaches seek to achieve their impacts. Nonetheless, these are aspects of delivery approaches that country research teams may investigate based on their relevance in their context.

Section 5 below discusses the DA functions and the conceptual distinction between the two main delivery pathways they can use. Section 6 then discusses the 'upstream' DA goals, inputs and context, and the 'downstream' intermediate and final outcomes. Section 7 then uses this framework to show how a range of different research and policy questions about DAs can be nested within this framework.

5. DA Functions and Pathways to Impact

In characterizing different delivery approaches, a key question is what managerial tools, levers, or strategies the approach seeks to deploy. We refer to these as delivery approach *functions*. We identify five sets of functions from which delivery approaches can draw on to achieve their objectives.[3] These are not only commonly observed across examples of delivery units but have also been identified in other delivery approaches in the policy and academic literature on implementation and governance, performance management, and public administration (e.g., World Bank, 2004; Laffont & Martimont, 2002; Hood, 1991; Locke & Latham, 1984). Different approaches may use and combine each in different ways, including not using some of them at all or using some of them inappropriately or ineffectively. Our purpose in listing them here is not to make any normative statements or theoretical hypotheses about their effectiveness but simply to categorize the range of options available to delivery approaches.

1) **Target setting and prioritization**. The establishment of a set of key priorities and objectives, measurable indicators to characterize progress against these

---

[3] As Table 1 illustrates, DAs can also perform other functions (e.g., training bureaucrats, transforming culture), so this set of five functions is meant to capture the core functions that DAs combine rather than to comprehensively enumerate the full universe of potential functions.

objectives, and benchmark levels of performance to be achieved in a specified time.

2) **Measurement and monitoring**. The establishment and execution of mechanisms to collect and report information about the performance of divisions, districts, teams, schools, and/or individuals across the organization or sector.

3) **Leveraging political sponsorship**. The leveraging and communication of high-level political backing for policy and service delivery. The audience for this signaling of political sponsorship of bureaucratic initiatives can be either the bureaucracy itself (to add pressure or legitimacy) or external stakeholders (to increase external pressure on the bureaucracy or serve as a commitment device for government to hold itself accountable). The attention and sponsorship of politicians itself are best understood as an input to the DA, but the DA's *leveraging* of this sponsorship through performance review routines and communication of political investment is a function of the DA.

4) **Accountability and incentives**. The establishment and execution of rewards and/or sanctions linked to performance – the 'carrots and sticks' associated with delivery approaches. This could include a range of types of incentives: monetary incentives, the threat of firing or other formal career incentives, reporting through high-stakes meetings which create strong reputational concerns, 'naming and shaming', or negative social perceptions.

5) **Problem-solving**. The routinization of mechanisms of dialogue, coordination, and problem-solving across multiple individuals, divisions, or organizations that can improve performance through better sharing of information and ideas. This could include horizontal collaboration and convening across teams, sectors, or actors, as well as the facilitation of 'bottom-up' approaches to catalyzing organizational learning through local problem-solving, adaptation, issue escalation, and policy feedback across the delivery chain.

These functions are not unique to delivery approaches; indeed, they describe the core of what most bureaucracies do on a day-to-day basis. A delivery approach might seek to improve performance by doing them differently or better or by combining them in unique ways. For example, most ministries have annual plans that define targets, but a delivery approach might combine the target-setting and prioritization process with leveraging political sponsorship to increase its salience, establish a higher-frequency measurement of performance, and establish routine performance review in the presence of the sector minister with naming-and-shaming of good and bad performers. But while this example illustrates a set of functions that are commonly bundled by many delivery approaches, the functions need not always be bundled. For instance, increased measurement of performance could be used for its informational value without combining it with grand targets or high-stakes accountability measures.

This list of the five main functions is intended to capture a set of functions performed by nearly all DAs, not to be exhaustive of all functions that can be performed by DAs. For example, many delivery approaches seek to provide training and capacity building as part of their efforts to support performance improvement (e.g., Barber et al., 2011; Scharff, 2012; Barber, 2015; Thomas, 2018). Other delivery approaches may conduct outreach to service beneficiaries and citizens to solicit feedback or build public pressure on service delivery agencies to improve performance (e.g., Barber, 2015; World Bank, 2017; McKay, 2017). Country research teams will seek to identify and explore these additional functions as relevant, and this empirical research can then inform the adaptation of this conceptual framework over the course of the DeliverEd project.

While this framework lays out the full range of forms and functions that delivery approaches can take on, within this it is possible to identify two distinct pathways or mechanisms that combine these functions and through which delivery approaches can have an impact. These pathways are not mutually exclusive, and in practice, most DAs combine them both in various ways and to differing extents, but it is useful to make a conceptual distinction between them to understand how and in which contexts delivery approaches influence service delivery quality. What we term 'Pathway A' relates to practices related to the creation of monitoring and accountability routines and the attachment of consequences and incentives to measured performance. The underlying theory behind this pathway is that policy implementation necessitates delegation to agents down the delivery chain who do not, a priori, share the same objectives as the policy-setting principal. It is assumed that agents are better informed about their actions and the context in which they operate than the principal and can use this information asymmetry to shirk or misdirect their effort. 'Pathway B', in contrast, encompasses practices that create routines for collaborative problem-solving and organizational learning. The underlying theory behind this pathway is that improving policy implementation requires agents to share in the ownership of goals and targets, collectively solve problems, coordinate with each other, and introduce local innovations to adapt policy to particular circumstances and circumvent constraints. Problem-solving of this sort requires agents to have or receive information and ideas about how best to act in particular circumstances, as well as authorization and routines to enable them to share and act on this information.

This conceptual distinction builds on and links to a longstanding body of theory and set of debates in public administration, economics, education, sociology, and management, relating to the foundations of bureaucratic motivation and behavior and thus how management practices should be adapted for different types of individuals, organizations, and tasks. On one hand, many authors focus on the potential for shirking and misalignment between the incentives of individual bureaucrats and the social good and thus emphasize the importance of tools of accountability, control, and incentives in driving effective bureaucracies (e.g., Finer 1941; Duflo et al., 2012; various in Finan et al., 2017; Leaver et al., in press). This approach has historically

been utilized both by rule-driven approaches to regulating bureaucratic behavior as well as by more recent and widespread reform trends that have sought to introduce stronger performance-linked incentives into public sectors worldwide (e.g., Dahlstrom & Lapuente 2010; Hasnain et al., 2012; Miller & Whitford, 2007; Perry et al., 2009). On the other hand, other authors emphasize bureaucrats' intrinsic motivation and professionalism and how providing discretion and autonomy and fostering bottom-up innovation and problem-solving can yield better results in some circumstances (e.g., Friedrich, 1940; Korten 1980; Simon, 1983; Argyis & Schön, 1997; Carpenter, 2001; Andrews et al., 2013; Andrews, 2015; Andersen & Moynihan, 2016; Honig, 2018; Rasul & Rogger 2018; Rasul et al., 2020). Within the education sector, this perspective is grounded in theories of distributed leadership (Coburn, 2016; Leithwood & Seashore-Louis, 2011), organizational learning (Coburn & Honig, 2008; Honig, 2008), and improvement science (Bryk et al., 2015; Huber, 1991), which draw from a long tradition in the learning sciences and study of human cognition and motivation that emphasize learning as a situated process that involves active creation of and utilization of knowledge in a group and organizational settings for deeper problem-solving (Lave & Wenger, 1991; Vygotsky, 1978). Similar insights into the importance of discretion and autonomy in the delivery of public services have been generated in research on international development, starting with Easterly's critique of top-down development and including efforts to support problem-driven adaptive iteration (Easterly, 2008; Woolcock & Pritchett, 2002). They have also generated new research on middle- and school-level management and its role in educational change (see for example Pritchett, 2015; World Bank, 2018; Levy et al., 2018). These two theories of behavior and management find perhaps their clearest expression in McGregor's classic (1960) distinction between 'Theory X', that workers are lazy and self-interested and need to be controlled and incentivized, and 'Theory Y', that workers are public-spirited professionals and should be supported to do their jobs as best they can — a distinction that has guided the development of much of modern management studies.

Of course, most authors recognize that both theories can pertain at the same time and to different extents in different contexts, and so a range of literature has discussed the types of contexts, tasks, individuals, and agencies for which different approaches might be more effective as well as the potential positive and negative interactions between them (e.g., Wilson, 1989; Le Grand, 2003; Chun & Rainey, 2005; Rasul & Rogger, 2018). This nuance also finds expression in literature on delivery units which discusses the ways in which these different tools can be drawn on and combined in the design and operation of delivery units (e.g., Andrews, 2014; Barber, 2015; Freeguard & Gold, 2015; CPI, 2016a; World Bank, 2017; Delivery Associates, 2018; Thomas, 2018). The recognition that both pathways co-exist in different forms and to different extents in most actually existing cases of delivery approaches is an important foundation for this conceptual framework and subsequent empirical analysis, as is the potential for positive and negative interactions between these conceptual strands.

Conceptually, Pathway A mainly utilizes four of the five functions discussed in the inventory approach. A political principal (e.g., Prime Minister or Minster) sets clear priorities which are embodied in measurable goals and targets (function 1), and measurement and monitoring routines are established and enacted to measure progress towards these targets (function 2). These indicators are periodically examined in high-stakes review forums and performance management routines, with rewards and sanctions being associated with the performance of individuals, organizations, or teams (function 4). High-level political sponsorship (function 3) underpins the legitimacy and effectiveness of each of these three steps; without this, targets would carry less weight, and the doling out of rewards and sanctions would struggle to be sustained in the face of inertia and opposition from vested interests. Like Pathway A, Pathway B utilizes target setting (function 1), measurement and monitoring, (function 2), and political sponsorship (function 3) as important tools for improving policy implementation. However, it also seeks to use these tools, not in service of accountability regimes for doling out rewards and sanctions, but to enhance the problem-solving and collaboration routines (function 5) that they seek to institutionalize. The intention is to enable iterative learning and adaptation through collaboration, problem-solving, and/or development of professional judgment — processes that change not only the individual behavior of agents but also social norms and organizational cultures for service delivery. In this way, targets and data are used to establish common understandings of what should be happening, what is happening, and what is preventing improved performance. Political sponsorship is used to support the prioritization of effort and common understandings of goals rather than the promise (and threat) of performance-linked incentives.

The UK PMDU and Malaysia's PEMANDU — two of the best-documented cases — provide useful illustrations of how these two pathways manifest themselves in actual cases, with each case combining both pathways in different ways. The UK's PMDU instituted a heavy regimen of stocktake meetings with the prime minister to monitor the cabinet's 17 priority policy areas, where the prime minister spent a substantive amount of time reviewing each Ministry's performance against pre-set targets (World Bank, 2010c; Barber, 2015; CPI, 2016a). While these regular stocktakes created high-stakes accountability, they were also leveraged to identify bottlenecks and brainstorm solutions with various stakeholders and officials at the highest level of government with the authority to potentially resolve high-level barriers to effective service delivery (such as budget constraints or inter-agency coordination, etc.) (World Bank, 2010c; CPI, 2016a; Gold, 2017). Quarterly meetings and other ad hoc updates were not only used as a forum to discuss ministers' performance against their targets but also to present progress reports on projects to identify binding constraints to bureaucrats' efforts and inform the rethinking of the implementation roadmap (World Bank, 2010c; CPI, 2016a; Barber, 2015).

Similarly, Malaysia's PEMANDU also combined Pathway A and Pathway B managerial levers, but in different ways. In Malaysia, policy labs were set up at the local level to

support the PEMANDU in the policy prioritization and target-setting process. The convening of six-week-long delivery labs was intended to bring together key stakeholders from different sectors and organizations who were collectively responsible for achieving key priorities (World Bank, 2010c; CPI, 2016b; World Bank, 2017; PEMANDU, n.d.). These labs were leveraged to identify initial targets, which were then reviewed, refined, and agreed upon between PEMANDU and Ministries, Departments, and Agencies (MDAs) (World Bank, 2010c; World Bank, 2017). A set of KPIs was assigned to each Minister in the form of a scorecard and a dashboard was developed to present data weekly to reflect the implementation progress of MDAs against their KPIs (World Bank, 2017). While the dashboard established incentives for ministers by demonstrating MDA performance in a relatively public manner, it was also used to uncover implementation problems in real-time and problem-solve as needed or escalate issues to the Prime Minister to unblock difficult challenges (CPI, 2016b; World Bank, 2017). As with the PMDU, this example illustrates how these two pathways can be drawn upon by different delivery approaches in different ways. A key focus of DeliverEd's empirical work will be to further explore how these two pathways were combined in each country case and to explore the ways in which these pathways did or did not produce their intended impacts on downstream bureaucratic knowledge, attitudes, behavior, policy implementation, and outcomes.

This conceptual framework is agnostic about the relative effectiveness of these two pathways and different ways of combining them. Instead, for research purposes, we view the choices made by policymakers in designing and implementing particular delivery approaches as *hypotheses* on the part of the designers about the nature of the underlying problem the approach is trying to solve, as well as about the relative effectiveness of different management tools for solving it. In this sense, the ways in which Pathways A and B (and the near-infinite set of ways in which they can be combined) are more effective in any given context is an empirical question that needs to be tested. The following section discusses the factors that might mediate or moderate their effectiveness in different contexts.

Our conceptual framework focuses primarily on identifying and distinguishing the range of design choices related to functions performed by a delivery approach, and it is not intended to represent a comprehensive set of factors that might influence how effective a DA is likely to be. In addition to the features, we highlight below as 'inputs' to the delivery approach (such as political sponsorship, clarity of goals, appropriate staffing), some internal features of DAs that are likely to affect its functioning include strong leadership, a highly motivated team, and performance-oriented organizational culture, and the careful management of relationships with other key stakeholders across and outside of government (e.g., Scharff, 2012; Barber et al., 2011; Barber, 2009; Andrews, 2014; Barber, 2015; Kohli, Moody, & Buskey, 2016; World Bank, 2017; O'Malley, 2019). For the sake of simplicity and flexibility, these and other potentially important factors are not highlighted explicitly in the theory of change or conceptual framework but will constitute important topics for empirical exploration in the country

cases, based on which appropriate changes may be made to the overall conceptual framework.

6. 'Upstream' and 'Downstream' Variables

DAs execute their functions with a given set of goals, inputs, and contextual features that are determined prior to, or 'upstream' from, the operation of the DA itself (as represented by the functions discussed in the previous section).[4] The operation of the DA then has impacts on a range of 'downstream' variables, some of which it can impact directly and others of which are more distal and can only be affected through these more proximate mediating variables of management and policy implementation. This section briefly discusses each set of variables and their role in the theory of change. It also gives some illustrative examples of key variables that fall within each category and makes some conceptual distinctions within them. However, the variables named are not exhaustive within each category, and empirical studies would have to stipulate which variables are most salient within each category in their particular case.

6.1 DA Goals
DAs are set up to achieve a range of goals that vary in their key characteristics. For conceptual purposes, we would consider these goals as the overarching service delivery goals of the political authority that motivated the creation of the DA. These are 'prior to' the operation of DAs in a logical sense and are conceptually distinct from the specific targets that the DA itself might set for the bureaucratic apparatus to achieve these targets. Of course, in practice, a goal might take the form of a specific politically defined target (e.g., to achieve a certain percentage improvement in test scores or to fully implement a policy by a certain date), and the DA itself might play an important role in formulating actionable plans or more specific targets, so this distinction is blurred in practice. As Figure 1 illustrates, there might also be potential feedback loops from the DA to its goals, although the extent of these is likely to vary across cases and designs.

The type of goals a DA seeks to achieve can potentially affect its operations and its impact. Some key distinctions that may be relevant include:

- **Time horizon of goals.** Whether the DA focuses primarily on short-term (1–2 years), medium-term (3–5 years), or long-term (6+ years) policies and goals.

- **Outputs versus outcomes.** Whether the DA aims to deliver a specific set of services, reforms, or interventions (outputs) or to achieve certain outcomes

---

[4] The idea of goals and inputs being prior to the operation of the DA is a conceptual distinction, not a temporal one. In practice, DAs can sometimes shape some of these goals, design decisions, and resource inputs themselves. We find the conceptual distinction useful nonetheless.

(e.g., improved test scores) without specifying the means through which these outcomes should be achieved.

- **Goal measurability**. The extent to which goals are tangible and easily measurable in their most important aspects.

- **Complexity:** The extent to which goals require numerous interlinked actions from many actors across the delivery chain.

- **Pre-specifiability of solutions.** The extent to which bureaucratic actions required to achieve desired outcomes can be fully specified in advance and in a uniform, top-down fashion. Goals whose solutions can be pre-specified are typical 'fidelity' problems, where the objective is to deliver a specific intervention with high fidelity. Goals whose solutions are difficult to pre-specify are typical 'adaptation' problems in which interventions cannot be fully specified in advance (and hence require bureaucratic discretion) or which need to be extensively adapted to particular contexts (and hence require local flexibility and adaptation).

- **Coordination.** The extent to which achieving goals requires action (and hence coordination) across multiple stakeholders, especially those outside government or who are not accountable to the DA's political sponsors.

This list merely illustrates some theoretical considerations for understanding how the initial goals of a DA might have consequences for variables in downstream categories (inputs, functions, management, outcomes). Section 7 below discusses how these theoretical distinctions can be used to generate research questions for empirical examination. Of course, over time the goals of a DA might evolve, potentially in response to feedback from the DA and its impacts, so understanding these longer-term evolutions is an important issue for empirical analysis.

6.2 DA Inputs

The inputs to a DA include high-level decisions about the design, resourcing, and governance of the DA. These inputs are what allow the DA to undertake its functions and may influence the effectiveness of these functions. These inputs can be understood as design decisions taken prior to the setting up of the DA, as well as decisions that continue to be made on an ongoing basis during the DA's existence (such as the annual allocation of resources through the budget or donor funding). While these design and resourcing decisions are logically prior to the operation of the DA in the sense that they are necessary for the DA to be able to exercise its functions, in practice there may of course exist feedback loops that allow the DA and its results to over time shape the inputs that it receives.

A (non-exhaustive) list of key inputs to DAs includes:

- **Integration with the bureaucracy.** Delivery approaches can be more or less integrated into existing, mainstream government bureaucracies in terms of structure, authority, and resourcing. In practice, a given delivery approach might combine multiple types of integration. Conceptually, we can distinguish two main approaches to structural integration:

    o **Using pre-existing structures:** New functions and responsibilities of the delivery approach are assigned to a pre-existing team or unit without a change to the overall organizational chart.

    o **Using new/re-organized structures:** Leverage some innovation in the organizational make-up of the agency to develop and operationalize the new delivery approach (for instance, setting up a new department or merging several departments and re-orienting their functions and tasks to accommodate the new delivery approach)

- **Level of operation**. While the archetypal delivery unit (and many subsequent approaches) may be based within the core executive (e.g., cabinet office, sector minister's office), delivery approaches can operate at a range of different levels of government. They can also operate across multiple levels – for instance by being housed in one level but with established liaisons at other levels. We categorize this into four sets of levels:

    o **Center-of-government:** The top-most level of political and bureaucratic authority, such as a president or prime minister's office.

    o **Ministerial:** A national-level sector ministry or other national-level agencies, such as an education ministry.

    o **Sub-national:** Any government entity that exists below the national level. This could be states, provinces, regions, districts, or municipalities.

    o **Frontline (e.g., school-level):** The level where service delivery takes place, i.e., where the front-line providers of the service directly engage with citizens. This would, for example, comprise schools in the education sector or basic health units in the health sector.

- **Financial resources**. The adequacy and reliability of financial resources to undertake DA functions, whether from the government budget or donors.

- **Staff and skills**. The number, quality, and fit of staff working within the DA. A key distinction is whether the approach is staffed by:

    o **Existing staff** operating within normal civil service roles and regulations.

    o **New/external staff** such as consultants, technical assistants, or other external hires working outside the mainstream civil service.

    o **A mixture** of existing and new staff.

- **Governance of the DA**. The oversight and autonomy of the DA, including the individuals or bodies to which it is accountable (both *de jure* and *de facto*), its powers and authority over other actors in the delivery chain (e.g., sector ministries, school districts), and the extent of its legal, financial, and policy autonomy. In low- and middle-income countries, these governance arrangements might include donor organizations as well as the government itself.

- **Political sponsorship**. The attention and support of high-level political figures for the DA, both initially and on an ongoing basis. While the extent of political sponsorship itself is best understood as an input to the DA (something that allows it to undertake its functions), the *leveraging* of this sponsorship by the technical staff of the DA is best understood as a function of the DA (something it does to influence the behavior of other actors in the delivery chain).

As with the DA goals, this list of inputs and design decisions is non-comprehensive as is intended to illustrate the range of variables that might be considered important inputs. These variables constitute potentially salient factors for consideration in future hypothesis formation, not hypotheses in themselves.

6.3 Political and Institutional Context

The range of contextual variables that can affect policy implementation and service delivery outcomes is nearly infinite. Rather than attempt to enumerate these, we instead focus on highlighting some contextual features that might interact with the operation of a DA and thus have implications for its design and effectiveness. While many such variables are idiosyncratic to particular contexts, several commonly important contextual features are:

- **Nature of political competition**. Countries vary in the extent to which the nature of political competition rewards quality of service delivery, allows for the government to effectively reward or sanction public personnel, and entails clientelism approaches to job patronage or public service delivery. These factors could either support or undermine the execution or effectiveness of accountability regimes or efforts to harness staff professionalism.

- **Time horizons**. Longer political and bureaucratic time horizons could make delivery approaches more effective, by making the promise/threat of reward/sanction in future periods more credible — or by enhancing staff's ability and desire to invest effort in collaborative efforts to share ideas and improve problem-solving.

- **Baseline management quality, routines, and performance**. Improving performance from a low level to a medium level might require a different approach (e.g., more top-down and accountability-driven) than improving

performance from a medium level to a high level (which might require more problem-solving and coordination) (Mourshed et al., 2007; Barber, 2009).

Other contextual or institutional factors that might be relevant for the operation of DAs in a given context could include the extent or character of decentralization, the nature of unions and public service bargains, or idiosyncratic features such as moments of crisis or the existence of recent or ongoing reforms in the country or sector.

6.4 Bureaucratic Functioning and Policy Delivery

The intended direct impacts of a delivery approach are to improve the functioning of the rest of the bureaucracy, in terms of the management practices and bureaucratic behaviors needed for effective policy implementation. This is captured in the theory of change in Figure 1 as entailing a change in bureaucratic functioning at various levels of the bureaucracy: national (e.g., the ministry of education), sub-national (e.g., a regional or district government), and at the level of the management of individual schools. These impacts might cascade down in a strictly hierarchical sense, where a DA located in the center of government shapes attitudes, behaviors, and actions within a sector ministry, which then shape the actions of sub-national bureaucracies, and so on down to the school level. Or, in the case of a DA which engages directly with multiple levels of government, the DA might have these effects directly on these subordinate levels — for example, when a DA sets district-level targets and links them to accountability measures, convenes problem-solving sessions that include school-level administrators, or aims to improve inter- or intra-sectoral coordination. These changes can also be intended to arise through other non-hierarchical means, for example by activating awareness and pressure from service beneficiaries onto frontline service delivery organizations or by providing better information from citizens to the center of government (although for simplicity these possibilities are not illustrated in the conceptual framework diagram).

The category 'bureaucratic functioning and policy delivery' refers broadly to changes in bureaucratic actions, behaviors, attitudes, management processes, and outputs that might result from the operation of a DA. This is a very wide category that itself captures a linked set of changes: a) delivery approaches introduce new management routines and activities, which in turn lead to b) changes in attitudes, skills, and behaviors of individual bureaucrats, which lead to c) improved policy delivery. Of course, given the complexities of organizational behavior and policy implementation, this set of changes is highly complex and the exact set of logical steps may vary across DA functions or policy areas, so this category would need to be elaborated to suit specific analytical contexts. While questions of empirical measurement are beyond the scope of this conceptual paper, for the sake of illustration this category could include everything from a generalized quantitative measure of organizational management quality for each district or school (e.g., Leaver et al., 2019) to nuanced qualitative data from interviews with key bureaucrats to the coding of administrative data on task or output

completion (e.g., Rasul et al., 2020). A detailed methodological approach to the study of changes in bureaucratic functioning and capacity will be developed to support country research teams' data collection and analysis.

A key conceptual distinction within this category is between bureaucratic actions and behaviors that are linked to *compliance* with mandated policies and processes as opposed to those that are linked to *organizational learning* that might result in better innovation, adaptation, and coordination. For example, implementing a pre-specified policy or plan with full fidelity may be relatively more compliance-related, while developing and implementing a school-based management improvement plan may be relatively more organizational learning-related — although almost all tasks clearly require both to at least some extent. This distinction is important because it relates to the distinction between the two pathways for DA functions, as Section 7 below discusses.

6.5 Final Outcomes

The ultimate goal of any delivery approach is to improve the final outcomes that matter to citizens, in terms of the quantity and quality of public service delivery. In the education sector, the most important such outcome is equitable and broad-based student learning. Another important (almost) final outcome for many education sector reform plans is teaching quality, which is closely linked to student learning (Hanushek & Rivkin, 2010; Hanushek et al., 2005; Rockoff, 2004; Sanders & Rivers, 1996; Snilstveit et al., 2015; Opper, 2019; Glewwe & Muralidharan, 2015). Outside of the education sector context, this could refer to the behavior of frontline bureaucrats more generally.

While improving these final outcomes is an important goal of any DA, the DA can only affect them through the mediating variables of management practices and policy implementation since frontline bureaucrats and students are too numerous for DAs to feasibly engage directly with. This distinction links back to the understanding of DAs as fundamentally being tools for improving management and policy implementation within the bureaucracy, with these improvements in bureaucratic functioning being thought to be important for (but distinct from) improving final service delivery outcomes. In this sense, a DA could narrowly succeed in improving the implementation of a policy by the bureaucracy without leading to an improvement in final outcomes, if the policy were the 'wrong' policy in the sense that it did not lead to the intended changes in outcomes. This distinction is important for evaluating the success of DAs, because there is a difference in root causes between a DA which fails to improve the implementation of a high-level policy directive and a DA that succeeds in improving policy implementation but fails to improve outcomes. After all, the high-level policy directive was the 'wrong' one for achieving its intended outcomes.

While the distinction between direct and ultimate outcomes of DAs risks reifying an overly rigid distinction between policy and implementation and is of course blurry in

practice, the conceptual distinction is important for guiding hypothesis formation and analyzing DAs' successes and failures. In practice, in the medium- to long-term DAs may have some degree of authority or involvement in setting policy, or in assessing policies' impact and feeding this back to high-level political or bureaucratic decision-makers. The extent to which they try and succeed to do this is an interesting question for empirical study.

7. Nesting Research Questions Within the Conceptual Framework

This conceptual framework can be used to generate research questions for further hypothesis formation and empirical study. In this concluding section, we aim to set out several sets of potential research questions that can be investigated, to illustrate how they can be nested within the overall conceptual framework. These broad sets of questions are not precise enough to constitute hypotheses in themselves, as are not rooted in specific theories of bureaucratic behavior or institutional forms, nor are they linked to specific empirical measures. Developing these more precise, testable hypotheses is left to future and ongoing work. The contribution of this section is rather to show how different types of questions (and associated hypotheses) can be nested within this conceptual framework, as this will guide how the disparate results of these subsequent empirical studies can be integrated into a more unified conceptual understanding of the design, operation, and impacts of delivery approaches.

The overall theory of change of DAs, as illustrated by Figure 1, is that the operation of a DA, informed by goals set for it and inputs given to it, will lead to an improvement in bureaucratic functioning and policy delivery that will (if the goals are the 'right' ones) lead to an improvement in final outcomes (*orange → blue → green → yellow*). However, as discussed above, we focus on the 'green' category of bureaucratic functioning and policy delivery (rather than on the 'yellow' final outcomes) as the key dependent variable because these are the factors that DAs seek to directly influence, with the effects of these direct impacts on the final outcomes being mainly a function of policy-setting rather than implementation and thus at least partially outside the control of the DA itself. And as Figure 1 illustrates, there may be important feedback loops from changes in bureaucratic functioning and policy delivery (green) back to the DA functions (orange), and from DA functions (orange) back to DA goals and inputs (blue).

This overall theory of change can usefully be broken up into smaller segments for studying its mechanisms. In particular, this conceptual framework lends itself to three sets of questions about how DAs might impact bureaucratic functioning: 1) the effects of DA functions on changing bureaucratic functioning (*blue → green*); 2) the effects of upstream factors on changing bureaucratic functioning, mediated through the effects of these upstream factors on the operation of the DA (*orange → blue → green*); and 3) how contextual factors interact with or moderate the impact of DA functions and/or DA goals and inputs on change in bureaucratic functioning (*gray x blue → green*, or

gray x orange → blue → green). The remainder of this section briefly discusses these three categories of research questions in turn.

For the first set of questions relating to the impact of DA functions on changes in bureaucratic functioning (*blue à green*), the overall question about DAs in their most general form is whether these delivery functions do, in fact, lead to changes in policy implementation. A more nuanced question is which bundle of functions (as represented by Pathways A and B, and their potential combinations) is most effective at achieving the desired improvements in bureaucratic functioning. And an even more nuanced set of questions would be about whether the two pathways are associated with different types of effects on downstream bureaucratic functioning variables – in particular, whether activities associated with Pathway A are relatively more likely to achieve improvements in compliance-oriented functioning and activities associated with Pathway B are more likely to influence organizational learning-related functioning. Answering these questions, captured in bullet points below, can help policymakers understand the overall effectiveness of DAs and potential trade-offs in their design.

- Does the creation and operation of a DA improve the functioning of the education sector bureaucracy?

- What are the impacts of accountability-oriented activities and problem-solving-oriented activities in improving the functioning of the education sector bureaucracy?

- Do accountability-oriented activities achieve relatively greater improvements in compliance-related functioning, and problem-solving-oriented approaches achieve relatively greater improvements in organizational learning-related functioning?

These questions are illustrative of the type of considerations that can fall within this category, not exhaustive or exclusive of other potential reformulations or more precise or context-specific hypotheses.

The second set of questions relates to the impacts of upstream factors (DA goals and DA inputs) involved in setting up, designing, and governing the DA on the changes in the functioning of the education bureaucracy, as mediated by the functioning of the DA itself (orange → blue → green). This captures the idea that some types of goals may be easier or harder to achieve with a DA (or with different types of DAs), and that the inputs given to DAs might affect the execution or effectiveness of the DA's functions in achieving improvements in downstream bureaucratic functioning. Importantly, the focus on the DA itself as mediating factor means that hypotheses in this category focus not on how upstream factors might themselves directly affect bureaucratic functioning (e.g. goals with pre-specifiable policy solutions are easier for bureaucracies to deliver, more resources improve bureaucratic functioning), but on how these upstream factors

affect the effectiveness of the DA in catalyzing these changes in bureaucratic functioning (e.g. having goals with pre-specifiable policy solutions makes pathway A relatively more effective, more resources improve the DA's ability to execute its functions).

- How do the goals of a DA (across different dimensions) affect its effectiveness in improving bureaucratic functioning?

- How do the inputs of a DA (across different dimensions) affect its effectiveness in improving bureaucratic functioning?

- How do the goals of or inputs to a DA (across different dimensions) influence the relative effectiveness of the two pathways in improving bureaucratic functioning?

Finally, the third set of questions are about the *interaction* of contextual factors with either DA functions or upstream DA goals and inputs in determining DA effectiveness (*gray x blue → green*, or *gray x orange → blue → green*). This captures the idea that some DA functions or pathways may be more or less effective in some contexts than others. Section 6.3 discusses some especially salient potential dimensions of context which could interact with these factors, and thus be used for hypothesis formation. Questions falling within this category could address:

- How does political and institutional context (across various dimensions) moderate the effectiveness of different DA functions, and of the two pathways?

- Does political and institutional context (across various dimensions) moderate the difficulty of achieving certain types of DA goals or the importance of certain types of DA inputs for a DA?

8. Conclusion

This paper's main contribution is the development of a conceptual framework for understanding and studying the effectiveness of delivery approaches in improving bureaucratic functioning and policy delivery. On its own, this conceptual framework does not provide any answers about whether delivery approaches are effective at achieving these objectives or how they can be most effective, but it does provide a structure for the future development of a set of research questions and analytical hypotheses that can be studied empirically across a range of contexts, and for integrating the findings from this disparate set of hypotheses and empirical studies into an updated understanding of delivery approaches and their effectiveness.

The subsequent empirical analysis will provide deeper and more precise insights into when, where, how, and why delivery approaches are best adopted (or avoided), but

pending this analysis, the conceptual framework itself can at least help academics and policymakers alike understand the right questions to ask about delivery approaches. The conceptual spread of delivery approaches has been nearly as rapid as their geographic spread, so we hope that this paper helps to provide some needed clarity on what delivery approaches are (and are not), the diversity of types of approaches, and potential trade-offs among them, and their relationship to other upstream, downstream, and contextual factors.

References

Alari, C., & Thomas, P. (2016). *Improving government effectiveness across the world.* Institute for Government. https://www.instituteforgovernment.org.uk/publications/improving-government-effectiveness-across-world

Alessandro, M., Lafuente, M., & Santiso, C. (2014). *Governing to Deliver: Reinventing the center of government in Latin America and the Caribbean*. Inter-American Development Bank. https://publications.iadb.org/en/governing-deliver-reinventing-center-government-latin-america-and-caribbean

Andersen, S. C., & Moynihan, D. P. (2016). Bureaucratic investments in expertise: Evidence from a randomized controlled field trial. *Journal of Politics, 78*(4),1032–1044.

Andrews, M., Pritchett L., & Woolcock M. (2013). Escaping capability traps through problem driven iterative adaptation (PDIA). *World Development, 51*, 234–244.

Andrews, L. (2014). *Ministering to education: A reformer's report.* Prathian Books.

Andrews, M. (2015). Doing complex reform through PDIA: Judicial sector change in Mozambique. *Public Administration and Development, 35*(4), 288–300.

Andrews, M., Pritchett, L., & Woolcock, M. (2017*). Building state capability: Evidence, analysis, action*. Oxford University Press.

Argyris, C., & Schön, D. A. (1997). Organizational learning: A theory of action perspective. *Reis, 77*(78), 345–348.

Aviv, R. (2014). Wrong answer: In an era of high-stakes testing, a struggling school made a shocking choice. *The New Yorker.* https://www.newyorker.com/magazine/2014/07/21/wrong-answer

Baars, S., Bernardes, E., Elwick, A., Malortie, A., McAleavy, T., McInerney, L., Menzies, L., & Riggall, A. (2014). *Lessons from London schools: Investigating the success.* Centre for London and CfBT Education Trust.

Barber, M. (2009). How government, professions and citizens combine to drive successful educational change. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins. (Eds.). *Second international handbook of educational change*. Springer*.*

Barber, M., Kihn, P., & Motfit, A. (2011, February). *Deliverology: From idea to implementation.* Our insights. McKinsey & Co. https://www.mckinsey.com/industries/public-sector/our-insights/deliverology-from-idea-to-implementation

Barber, M. (2015). *How to run a government: So that citizens benefit and taxpayers don't go crazy.* Penguins.

Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration, 84*(3), 517–538.

Bloom, N., & Reenen, J. V. (2007). Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics, 122,* 1351–1408.

Bloom, N., Sadun, R., & Reenen, J. V. (2012). "The Organization of Firms Across Countries." *Quarterly Journal of Economics, 127*, 1663–1705.

Bloom, N., Renata L., Sadun, R., & Reenen, J. V. (2014). Does management matter in schools. *NBER Working Paper, 20667*. National Bureau of Economic Research.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.

Carpenter, D. (2001). *The forging of bureaucratic autonomy: Reputations, networks, and policy innovation in executive agencies, 1862–1928*. Princeton University Press.

Centre for Public Impact (CPI). (2016a). *The Prime Minister's Delivery Unit (PMDU) in the UK*. Boston Consulting Group. https://www.centreforpublicimpact.org/case-study/prime-ministers-delivery-unit-uk/

Centre for Public Impact (CPI). (2016b). *Malaysia's Performance Management and Delivery Unit.* Boston Consulting Group. https://www.centreforpublicimpact.org/case-study/performance-management-delivery-unit-kuala-lampur/

Chambers, V., Wild, L., King, M., & Harris, D. (2012). *Common constraints and incentive problems in service delivery* (Working Paper 351). ODI.

Chun, Y. H., & Rainey, H. G. (2005). Goal ambiguity and organizational performance in U.S. federal agencies. *Journal of Public Administration Research and Theory, 15*, 529–557.

Coburn, C. E. (2016). What's policy got to do with it? How the structure-agency debate can illuminate policy implementation. *American Journal of Education, 12*(3), 465–475.

Daly, E., Riese, J., & Singham, S. (2015). *The delivery challenge: A systematic approach to achieving breakthrough impact.* McKinsey & Company. https://www.mckinsey.com/industries/public-sector/our-insights/the-delivery-challenge-a-systematic-approach-to-achieving-breakthrough-impact

Dahlstrom, C., & Lapuente, V. (2010). Explaining cross-country differences in performance-related pay in the public sector. *Journal of Public Administration Research and Theory, 20*(3), 577–600.

Delivery Associates. (2018). *Success delivered: How delivery units make a difference for governments and the citizens they serve.* https://www.deliveryassociates.com/publications/success-delivered

Dumas, V, Lafuente, M., & Parrado, S. (2013). *Strengthening the center of government for results in Chile: The experience of the ministry of the presidency and its President's Delivery Unit (2010–2013).* Inter-American Development Bank. https://publications.iadb.org/en/strengthening-center-government-results-chile-experience-ministry-presidency-and-its-presidents

Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives work: Getting teachers to come to School. *American Economic Review, 102*(4),1241–1278.

Easterly, W. (2008). Institutions: Top down or bottom up? *American Economic Review, 98*(2), 95–99.

Finan, F., Olken, B. A., & Pande, R. (2017). The personnel economics of the state. In A. V. Banerjee, & E. Duflo Elsevier (Eds.), *Handbook of economic field experiments* (pp. 467–514). North-Holland.

Freeguard, G., & Gold, J. (2015). *Data-driven delivery: lessons from the O'Malley Administration of Maryland.* Institute for Government. https://www.instituteforgovernment.org.uk/publications/data-driven-delivery-lessons-o'malley-administration-maryland

Friedrich, C. (1940). Public policy and the nature of administrative responsibility. *Public Policy, 1*, 3–24. Reprinted in F. Rourke (Ed.) *Bureaucratic power in national politics* (3rd ed., 1978). Little, Brown.

Finer, H. (1941). Administrative responsibility in democratic government. *Public Administration Review, 1*(4), 335–350.

Gibbons, R., & Henderson, R. (2013). What do managers do? In R. Gibbons, & J. Roberts (Eds.), *The handbook of organizational economics*. Princeton University Press.

Gold, J. (2014). *International delivery: Centres of government and the drive for better policy implementation.* Institute for Government. https://www.instituteforgovernment.org.uk/sites/default/files/publications/International%20Delivery%20report.pdf

Gold, J. (2017). *Tracking delivery: Global trends and warning signs in delivery units.* Institute for Government. https://www.instituteforgovernment.org.uk/sites/default/files/publications/Global%20Delivery%20report.pdf

Glewwe, P., & Muralidharan, K. (2015). *Improving school education outcomes in developing countries: Evidence, knowledge gaps, and policy implications.* RISE. https://www.riseprogramme.org/sites/www.riseprogramme.org/files/publications/RISE_WP-001_Glewwe_Muralidharan.pdf

Grindle, M. (1997). Divergent cultures? When public organizations perform well in developing countries. *World Development, 25*, 481–95.

Gruening, G. (2001). Origin and theoretical basis of new public management. *International Public Management Journal*, 4(1), 1–25.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (No. w11154). National Bureau of Economic Research.

Hanushek, E. A., & Rivkin, S. G. (2010). The quality and distribution of teachers under the No Child Left Behind Act. *Journal of Economic Perspectives, 24*(3), 133–150.

Harrison, T. (2016, July). *The role of the centre in driving government priorities: The experience of delivery units* (Working Paper). Oxford Policy Management. https://www.opml.co.uk/files/Publications/corporate-publications/working-papers/wp-role-centre-driving-government-priorities.pdf?noredirect=1

Hart, C. (2017). *The Albania Delivery Unit: A case study on accountability in action* (Working Paper). https://albania.growthlab.cid.harvard.edu/files/albaniagrowthlab/files/hart_2017_accountability_in_action_-_albanian_delivery_unit.pdf?m=1527708899

Hasnain, Z., Manning, N., & Pierskalla, J. H. (2012). *Performance- Related Pay in the Public Sector* (Policy Research Working Paper 6043). World Bank. http://documents.worldbank.org/curated/en/666871468176639302/Performance-related-pay-in-the-public-sector-a-review-of-theory-and-evidence

Honig, M. I. (2008). District central offices as learning organizations: How sociocultural and organizational learning theories elaborate district central office administrators' participation in teaching and learning improvement efforts. *American Journal of Education 114*(4), 627–664.

Honig, M. I., & Coburn, C. (2008). Evidence-based decision making in school district central offices: Toward a policy and research agenda. *Educational Policy, 22*(4), 578–608.

Honig, D. (2018). *Navigation by judgment: Why and when top-down management of foreign aid doesn't work.* Oxford University Press.

Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. *Organization Science*, *2*(1), 88–115.

Hood, C. (1991). A public management for all seasons? *Public Administration, 69*(1), 3–19.

Hood, C., & Dixon, R. (2010). The political payoff from performance target systems: No-brainer or no-gainer? *Journal of Public Administration Research and Theory*, *20*(2), 281–298.

Kohli, J., & Moody, C. (2016). *What is a delivery unit?* (Delivering results series). Deloitte. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/us-fed-what-is-a-delivery-unit.pdf

Kohli, J., Moody, C., & Buskey, M. (2016). How should a delivery unit be designed? (Delivering results series). Deloitte. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/us-fed-delivering-results-series.pdf

Korten, D. (1980). Community organization and rural development: A learning process approach. *Public Administration Review, 40*(5), 480–511.

Lafuente, M., & Gonzalez, S. (2018). *Do delivery units deliver?: Assessing government innovations* (Technical Note No IDB-TN-1431). Inter-American Development Bank*,* Innovations in Citizen Services Division.

Laffont, J., & Martimort, D. (2002). *The theory of incentives: The principal-agent model.* Princeton, N.J.

Langford, J., & Roy, J. (2008). Integrating service delivery across levels of government: Case Studies of Canada and other countries. IBM Center for the Business of Government.

Lave, J., & Wenger, E. (1991)*. Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Leaver, C., Lemos, R., & Scur, D. (2019). *Measuring and explaining management in schools: New approaches using public data.* RISE. https://assets.publishing.service.gov.uk/media/5ed0e8f5d3bf7f45fcf2c5b2/RISE_ WP-033_Leaver_Lemos_Scur.pdf

Leaver, C., Ozier, O., Serneels, P., & Zeitlin, A. (in press). Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools. *American Economic Review.*

Le Grand, J. (2003). *Motivation, agency, and public policy: Of knights and knaves, pawns and queens*. Oxford University Press.

Leithwood, K., & Seashore-Louis, K. (2011). *Linking leadership to student learning*. John Wiley & Sons.

Levy, B., & Kpundeh, S. J. (2004). *Building state capacity in Africa: New approaches, emerging lessons.* World Bank Publications.

Levy, B., Cameron, R. Hoadley, U., & Naidoo, V. (Eds.). (2018). *The politics and governance of basic education: A tale of two South African provinces.* Oxford University Press.

Lindquist, E. (2006). Organising for policy implementation: The emergence and role of implementation units in policy design and oversight. *Journal of Comparative Policy Analysis: Research and Practice, 8*(4), 311–324.

Lindquist, E. A., & Wanna, J., (2010). Delivering policy reform: making it happen, making it stick. In *Delivering policy reform: Anchoring significant reforms in turbulent times.* Griffith University.

Locke, E. A., & Latham, G. P. (1984). *Goal setting: A motivational technique that works!* Prentice Hall.

McDonnell, E. (2017). Patchwork leviathan: How pockets of bureaucratic governance flourish within institutionally diverse developing states. *American Sociological Review, 82*(3), 476–510.

McDonnell, E. (2020). *Patchwork leviathan: Pockets of bureaucratic effectiveness in developing states*. Princeton University Press.

McKay, S. (2017, March 2). *Quality improvement: Deliverology.* Carnegie Foundation for the Advancement of Teaching. https://www.carnegiefoundation.org/blog/quality-improvement-deliverology/

McGregor, D. (1960). *The human side of enterprise.* McGraw-Hill.

Miller, G., & Whitford, A. B. (2007). The principal's moral hazard: Constraints on the use of incentives in hierarchy. *Journal of Public Administration Research and Theory, 17*(2), 213–233.

Milward, B. (1982). Interorganizational policy systems and research on public organizations. *Administration and Society, 13*(4), 457–478.

Mourshed, M., Chijoke, C., & Barber, M. (2007). How the school's most improved school systems keep getting better. McKinsey & Company. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/how-the-worlds-most-improved-school-systems-keep-getting-better

National Committee for Human Resource Development (NCHRD). (2016). *National strategy for human resource development 2016–2025*. https://www.hrd.jo/nationalstrategy

Nordstrum, L. E., LeMahieu, Paul G., & Dodd, K. (2017). Deliverology in Kentucky. *Quality Assurance in Education*, *25*(1), 43–45.

O'Malley, M. (2019). *Smarter government: How to govern for results in the information age.* Esri Press.

Opper, I. M. (2019). *Teachers matter: Understanding teachers' impact on student achievement.* RAND Corporation. https://www.rand.org/pubs/research_reports/RR4312.html

PEMANDU Associates. (n.d.). *Big fast results (BFR) Methodology: 8 steps of transformation.* https://pemandu.org/8-steps-of-transformation/

Perry, J., Engbers, T., & Jun, S. Y. (2009). Back to the future? Performance-related pay, empirical research, and the perils of persistence. *Public Administration Review, 69*(1), 39–51.

Pressman, J., & Wildavsky, A. (1979). *Implementation: How great expectations in Washington Are dashed in Oakland: Or, why it's amazing that federal programs work at all, this being a saga of the Economic Development Administration as told by two sympathetic observers who seek to build morals on a foundation of ruined hopes.* University of California Press.

Pritchett, L. (2013). *The rebirth of education: Schooling ain't learning.* Center for Global Development Books.

Pritchett, L. (2015). *Creating education systems coherent for learning outcomes* (Working Paper 15/005). RISE. www.riseprogramme.org/publications/rise-working-paper-15005-creating-education-systems-coherent-learning-outcomes

Rasul, I., & Rogger, D. (2018). Management of bureaucrats and public service delivery: Evidence from the Nigerian civil service. *Economic Journal, 128*, 413–146.

Rasul, R., Rogger, D., & Williams, M. J. (2020). Management, organizational performance, and task clarity: Evidence from Ghana's civil service. *Journal of Public Administration Research and Theory*.

Rainey, H. G., & Steinbauer, P. (1999). Galloping elephants: Developing elements of a theory of effective government organizations. *Journal of Public Administration Research and Theory, 9,*1–32*.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247–252.

Rose-Ackerman, S. (1986). Reforming public bureaucracy through economic incentives? *Journal of Law, Economics and Organization, 2,*131–61.

Richards, D., & Smith, M. (2006). Central control and policy implementation in the UK: A case study of the Prime Minister's Delivery Unit. *Journal of Comparative Policy Analysis*, *8*(4), 325–345.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* University of Tennessee Value-Added Research and Assessment Center.

Scism, D. (2015). *The New 'Deliverology': 14 years of evolution*. London School of Economics and Centre for Public Impact.

Scharff, M. (2012). *Delivering on a presidential agenda: Sierra Leone's Strategy and Policy Unit, 2010–2011.* Innovation for Successful Societies, Princeton University.

Scharff, M. (2013*). Translating vision into action: Indonesia's Delivery Unit, 2009– 2012*. Innovations for Successful Societies*,* Princeton University.

Simon, W. (1983). Legality, bureaucracy, and class in the welfare system. *Yale Law Journal, 92*, 1198–1269.

Simson, R. (2013). *Unblocking results: Case Study – The Africa Governance Initiative in Sierra Leone.* Centre for Aid and Public Expenditure, ODI.

Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T., & Eyers, J. (2015, December). *Interventions for improving learning outcomes and access to education in low-and middle-income countries: A systematic review* (*3ie final review*.) International Initiative for Impact Evaluation.

Shostak, R., J. Watkins, Bellver, A., & John-Abraham, I. (2014). *When might the introduction of a delivery unit be the right intervention?* World Bank Governance and Public Sector Management Practice Notes.

Thomas, L. (2018). *Prime Minister's Delivery Unit will be data driven.* Bahamas Information Services. http://www.thebahamasweekly.com/publish/bis-news-updates/Prime_Minister_s_Delivery_Unit_will_be_data_driven58031.shtml

Tiernan, A. (2006). Working with the stock we have: The evolving role of Queensland's Implementation Unit. *Journal of Comparative Policy Analysis: Research and Practice, 8*(4), 371–391.

Todd, R. (2014). *Delivery units: Can they catalyze sustained improvements in education service delivery?* Cambridge Education. http://devpolicy.org/2014-Australasian-Aid-and-International-Development-Policy-Workshop/Papers/4b/Andy-Brock.pdf

Todd, R., & Waistell, D. (in press). *Overview of the delivery approach* (Report for the Bill and Melinda Gates Foundation). Cambridge Education.

Vygotsky, L. S. (1978). *Mind in society.* Harvard University Press.

Wilson, J. (1989). *Bureaucracy: What government agencies do and why they do it.* Basic Books.

Woolcock, M., & Prichett, L. (2002). *Solutions when the solution is the problem: Arraying the disarray in development* (Working Paper No. 10). Center for Global Development.

World Bank. (2004). World development report: World Bank world development Report 2004. *Development and Change, 36*(6), 1263–1266. https://openknowledge.worldbank.org/handle/10986/5986

World Bank. (2010a). *Improving performance: Foundations of systemic performance* (Brief). Global Expert Team (GET). http://documents.worldbank.org/curated/en/923311468337217269/pdf/600920BRI0GET010BOX358310B01PUBLIC1.pdf

World Bank. (2010b). *Ghana's new approach to public sector reform: Focusing on delivery* (Brief No. 70156). Global Expert Team (GET). https://openknowledge.worldbank.org/bitstream/handle/10986/12811/701560ESW0P1230Delivery0Model0Note.pdf?sequence=1&isAllowed=y

World Bank. (2010c). *Driving performance through the center of government delivery units* (Brief). Global Expert Team (GET). http://documents1.worldbank.org/curated/en/923311468337217269/pdf/600920BRI0GET010BOX358310B01PUBLIC1.pdf

World Bank. (2017). *Driving performance from the center: Malaysia's experience with PEMANDU (*English). http://documents.worldbank.org/curated/en/318041492513503891/Driving-performance-from-the-center-Malaysia-s-experience-with-PEMANDU

World Bank. (2018). *World development report: Learning to realize education's promise.* https://www.worldbank.org/en/publication/wdr2018.