the Education Commission

BLAVATNIK SCHOOL OF GOVERNMENT

UNIVERSITY OF OXFORD

UKaid from the British people

Working Paper

DeliverEd: Research on How to Deliver Education Reforms

Delivery Approaches to Improving Policy Implementation: A Conceptual Framework

Version: July 8, 2020

Martin J. Williams, Clare Leaver, Zahra Mansoor, Dana Qarout, and Anna Bilous, Blavatnik School of Government, University of Oxford

Karen Mundy, Minahil Asim, Sheena Bell, University of Toronto

The DeliverEd initiative was established in response to a recommendation from the Education Commission's Learning Generation report to explore new ways of achieving results towards learning for all. The Education Commission and Oxford University's Blavatnik School of Government are partnering to lead this initiative, with support from the University of Toronto, as well as leading local research organizations in five countries. DeliverEd is supported by the UK's Department for International Development.

For other supporting documents, please visit https://educationcommission.org/delivered-initiative/.

This working paper focuses on how to define delivery and a conceptual framework for delivery approaches in education. For questions or more information about this paper, please contact: info@educationcommission.org.

Please reference this publication as: Williams, Martin J., Clare Leaver, Zahra Mansoor, Dana Qarout, and Anna Bilous (University of Oxford), Karen Mundy, Minahil Asim, and Sheena Bell (University of Tornoto). 2020. "Delivery Approaches to Improving Policy Implementation: A Conceptual Framework." Working Paper for the DeliverEd Initiative. New York: Education Commission.

1. Introduction

Governments all around the world are adopting ambitious goals and reforms aimed at improving the quality of service delivery, but often struggle to translate these reform intentions into tangible action because these high-level intentions have to be carefully formulated into an actionable plan, transmitted through a complex and multi-sited bureaucratic apparatus, and then put into practice by frontline bureaucrats. However, achieving rapid changes in the functioning of service delivery bureaucracies can be challenging, both due to potential bureaucratic inertia and resistance as well as to the complexity, coordination, discretion, and innovation that can be 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. Widely viewed as having started with the UK's Prime Minister's Delivery Unit (PMDU) in the early 2000s, these delivery approaches have been adapted and adopted in various forms by dozens of governments worldwide. The primary objective of these delivery approaches is to improve policy delivery by changing bureaucratic behavior. But while their status as a 'go-to' tool for efforts to improve bureaucratic functioning is evident from their rapid spread and popularity with government leaders and donors alike, they have sometimes been touted as a 'cure-all' for a broad range of goals and problems, apparently applicable to all contexts and all purposes.

But as with any management tool or institutional form, delivery approaches can be designed in different ways. These design choices 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. More broadly, such a conceptual framework is necessary to understand how delivery approaches relate to other approaches to improving bureaucratic functioning and policy delivery in terms of potential alternatives or complements to delivery approaches, but also in terms of what existing structures and processes exist in a given system into which a delivery approach is being introduced.

This paper puts forward a conceptual framework for understanding delivery approaches: what they are and seek to do, different ways in which they can be designed, and the trade-offs that these design choices entail. 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 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 our minister might choose, and which type of approach 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 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 improving the performance of the education sector as a whole. Our 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. We illustrate the potential connections and feedback loops among these categories.

Section 5 discusses the range of functions undertaken by delivery approaches in more detail, classifying them into five categories: target setting, monitoring and measurement, leveraging political sponsorship, accountability and incentives, and problem solving. While most delivery approaches draw on all five to varying degrees, we identify two ideal types1 that represent alternative theories of change on how to improve bureaucratic functioning and policy delivery. One style of delivery approach relies heavily on top-down tools to increase effort among bureaucrats and school management, in particular through accountability regimes and incentives (financial or non-financial). We term this style 'Type A'. Another style is to rely more on bottom-up ways to stimulate organizational learning, coordination, and problem-solving among bureaucrats and school management. We term this style 'Type B'. We describe these two types in more detail, and relate them to observed empirical cases. While in practice most actual delivery approaches combine both styles to some degree, and there are several functions that are in common across both approaches, the conceptual distinction

1 These are 'ideal types' in the sense that they represent hypothetical conceptually pure versions of the underlying concept, but should not be understood as 'ideal' in the normative sense that policymakers should aim to actually operationalize them in their pure form.

is useful because these two ideal types represent different understandings and hypotheses about the underlying causes of poor policy implementation. One contribution of our conceptual framework is to make these alternative designs, associated theories of change, and links to goals and context more explicit, in order to enable empirical study by researchers and informed decisions by policymakers.

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, and 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 analyzed empirically, because this would require adopting particular 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 8 concludes.

While our conceptual framework is applicable to delivery approaches at all levels and sectors, we focus our discussion in this paper primarily 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.

This conceptual paper provides the foundation for more detailed methodological development of a research design for studying the effects of different types of delivery approaches across several country cases. This work will build methodologically on these conceptual foundations to operationalize these core concepts and hypotheses. This research design will then be adapted and contextualized to the country cases. Taken together, this research will provide rigorous evidence on the uses and effectiveness of a range of delivery approaches.

Within this, the primary purpose of this paper is to provide a framework for the future formulation and empirical examination of hypotheses about delivery approaches and their effectiveness, and the findings of such analysis will of course be of direct relevance to policymakers. However, the conceptual framework itself is also likely to be useful to policymakers who are considering adopting a delivery approach or are already involved in one. Existing experience has shown that delivery approaches are often adopted without a clear understanding of why particular management arrangements are adopted or precisely what problem they aim to solve. By clarifying the conceptual space around delivery approaches and reviewing what is and is not known about their effectiveness, this paper may help policymakers understand what a delivery approach is and the different varieties of delivery approaches, what a delivery approach might aim to achieve (and what might be beyond its scope), and on what types of factors its effectiveness could depend.

2. Existing Theory and Evidence on Delivery Approaches

2.1 History and Prior Conceptualizations

The most widely known type of delivery approach is the delivery unit. The first model of the delivery unit emerged in the UK in 2001 (Prime Minister's Delivery Unit – PMDU) which drew inspiration by the model of performance management developed by the New York Police Department in the 1990s (Gold, 2017). In their purest form, these take the form of a new unit located in an executive office (e.g. president's office, minister's office) and staffed by technical analysts drawn largely from outside the core civil service. This unit's function is to amplify the technical oversight capacities of the executive leader in order to intensify and routinize processes of target-setting, measurement and monitoring of performance, and accountability. The early PMDU was characterized mainly by a top-down, accountability-driven, 'targets and terror' approach (Bevan & Hood, 2006). Delivery units are often closely linked – in perception, approach, and advisory or contractual relationships – to Michael Barber and the consulting firms Delivery Associates and McKinsey, which helped export many of the ideas behind the UK PMDU from the mid-2000s under the label of 'deliverology' (Barber et al, 2011, Kohli et al, 2016). 'Deliverology' is based on three key principles: setting up a delivery unit, defining targets, and instituting data collection mechanisms (Barber et al, 2011).

Many delivery units created based on the model of the UK PMDU shared commonalities with it, such as the Punjab Roadmap in Punjab, Pakistan, or the Premier's Implementation Unit in New South Wales in Australia, involving centralized target setting with a small team along with strong involvement of senior leadership. However, variations in the form of delivery approaches have also emerged (Todd & Waistell, forthcoming). For example, some delivery units (such as Malaysia's PEMANDU) conduct 'delivery labs' (intensive 6-9-week problem-solving sessions) with key stakeholders to identify and resolve problems. These approaches share with delivery units a reliance on high-profile targets and data analysis, but focus their efforts more on cross-sectoral convening of stakeholders and problem-solving rather than on enacting rewards or sanctions (Daly et al, 2015; Daly & Singham, 2012; World Bank, 2017). Other countries, such as Denmark, have adopted delivery labs as part of a service delivery improvement strategy without an associated delivery unit (Buerkli, 2015).

Other cases have moved even farther from the PMDU's top-down, accountability-focused approach. For example, the Kentucky Commission Delivery Unit (CDU) at the Kentucky Department of Education in the US sought to achieve improved results through distributed decision-making practices (Nordstrum et al, 2017). The CDU aimed to align the efforts of central and district level authorities in improving access to professional education. It developed an action research agenda including research questions, strategy and specific indicators

through a consultative process around defining questions that needed to be answered, specific on career and college readiness, and the definition of these. These were converted into delivery plans which facilitated the cross-functional collaboration within and between department and school districts. As with most delivery units, the CDU used a set of performance management routines to discuss data regularly, but used these not for accountability purposes but to arrive at a shared view of progress and collectively make decisions (Nordstrum et al, 2017).

Several authors have attempted to distil the key elements, principles, or functions of delivery approaches into typologies. Table 1 presents a selection of prominent classifications. While there are variations across these definitions, several common functions emerge. These include: a focus on setting ambitious targets and priorities for service delivery improvement; a heavy reliance on data-driven decision-making, often informed by improved data collection and review routines; and direct involvement from high-level political sponsors. Most discussions also emphasize the importance of staffing delivery units with talented personnel; while this is often implicitly interpreted to mean bringing in staff or consultants from outside the civil service, this is not always so (Gold, 2017; Shostak et al, 2014; Barber et al, 2011). Similarly, many descriptions of delivery units explicitly or implicitly assume they should be situated at the center of government and interact mainly with other high-level national policymakers, but in practice many delivery approaches sit entirely or partially within either sectoral ministries or sub-national levels of government (Gold, 2017).

As Table 1 indicates, there is no universally agreed-on definition of delivery units (or of the broader concept of delivery approaches), and indeed existing definitions tend to be constructed based on normative prescriptions for what they should look like rather than describing their conceptual boundaries or distinguishing among different types of approaches. Some definitions of delivery approaches, such as Gold's (2014) and Shostak et al's (2014), describe delivery units as one tool used by Centre of Government to achieve a limited number of outcomes that are top priority for government rather than a

Table 1: Selected Typologies of Delivery Unit Functions[2]

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 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. Solicit feedback from clients (citizens/beneficiaries) and other stakeholders to revise implementation plans 4. Communicate progress regularly and widely to signal transparency and accountability for results (including to the public) 5. Set targets and review them periodically based on analysis of data 6. Establish monitoring and problem-solving routines 7. Validate and verify 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 that sufficient resources are allocated for implementation of these priorities. 2. Setting up data collection, analysis and information sharing routines to inform decisions. 3. Analyzing delivery challenges by engaging with various stakeholders and frontline practitioners (to understand incentives/perception of agents) 4. Maintaining open communication channels with delivery agents and the public, and ensuring accountability relationships with responsible stakeholders/agents 5. Securing and signaling senior leadership support and commitment to priorities through communication and	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 Centre 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 Typologies listed by author

routine monitoring and
reporting practices.

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.

2.2 Empirical Studies

There exists a plethora of policy literature documenting claimed successes and challenges of delivery units worldwide, but much of this literature is based on before-vs-after comparisons of indicators or on 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.

Nonetheless, many delivery units have been associated with significant improvements in some indicators. 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. In Colombia, theft of cell phones declined by 12% in a single year. 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 US 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 the process of 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, 2019).

Some studies have pointed out potential evolutions in the character of delivery approaches over time. For example, Scism et al (2015) discusses the expansion of functions beyond monitoring and performance

management towards less top-down 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 Albania, Tanzania, Liberia, Canada, Brazil, and the UK (Gold, 2017; Roy and Langford, 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 and 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).3

The policy-oriented and limited academic literature that exist on delivery approaches also highlight some common challenges. These include: a) dealing with political transitions (Barber et al 2011; Freeguard and Gold 2015, NCHRD 2016); b) limited influence over delivery chains (Freeguard and Gold 2015; Delivery Associates 2018); c) entrenched operating cultures and practices (Barber et al 2011; Gold 2017); d) measuring relevant outcomes (Barber et al 2011; Aviv 2014; Freeguard and Gold 2015); and e) citizen engagement to maintain political support and solicit inputs (Freeguard and Gold, 2015; Shostak et al, 2014; DA, 2018; CPI 2016a).

3. Definition and Scope of Delivery Approaches

We build on this existing literature by defining 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. 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

3 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

distinctions also serve a substantive purpose, as they help to clarify what delivery approaches do and what alternative approaches for achieving similar objectives exist.

While this section mainly focuses on the boundaries of delivery approaches, we begin by briefly discussing different types of delivery approaches, in particular the relationship of our concept of delivery approaches to delivery units. We view delivery units as one type of delivery approach, which tend to be located at the center of government and have often (but not always) focused on top-down approaches to target-setting, measurement, and accountability. This contrasts with the delivery lab approach, which focuses more on problem-solving and convening of stakeholders from across sectors and/or across the delivery chain, especially at the agenda-setting stage. Of course, these distinctions are blurred in both theory and practice, as improving cross-sectoral collaboration is also identified as an important element of 'deliverology' by its proponents, and top-down accountability through delivery units can form a part of the delivery lab approach (as in Malaysia's PEMANDU). Furthermore, our conceptual distinction between 'unit' and 'lab' does not necessarily correspond to the names of actual units and labs, and there exist different approaches among units and labs as well as hybrid approaches that combine both (such Malaysia's education sector Performance and Delivery Unit). An additional complication is that what delivery units or labs claim to do in their rhetoric is not necessarily what they actually do in practice – it is easy to claim that a given approach combines the benefits of both accountability-driven and problem-solving-driven approaches, but the extent and nature of trade-offs that exist is an important theoretical and empirical question.

Furthermore, the use of data, targets, accountability, or problem-solving routines was neither invented by nor is restricted to delivery units and labs. Indeed, these are core functions of all bureaucracies, and delivery units and labs often have very 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 directorates, etc.). 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 discuss these cases in more depth in section 4.

Second, we identify delivery approaches as those that are spearheaded by government. Initiatives that are exclusively externally led by NGOs, donors, or local communities are not best understood as delivery approaches.4 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, which 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 the 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 with the intention of being temporary in nature, for example for use in responding to a specific crisis or enacting a broad program of bureaucratic reforms.

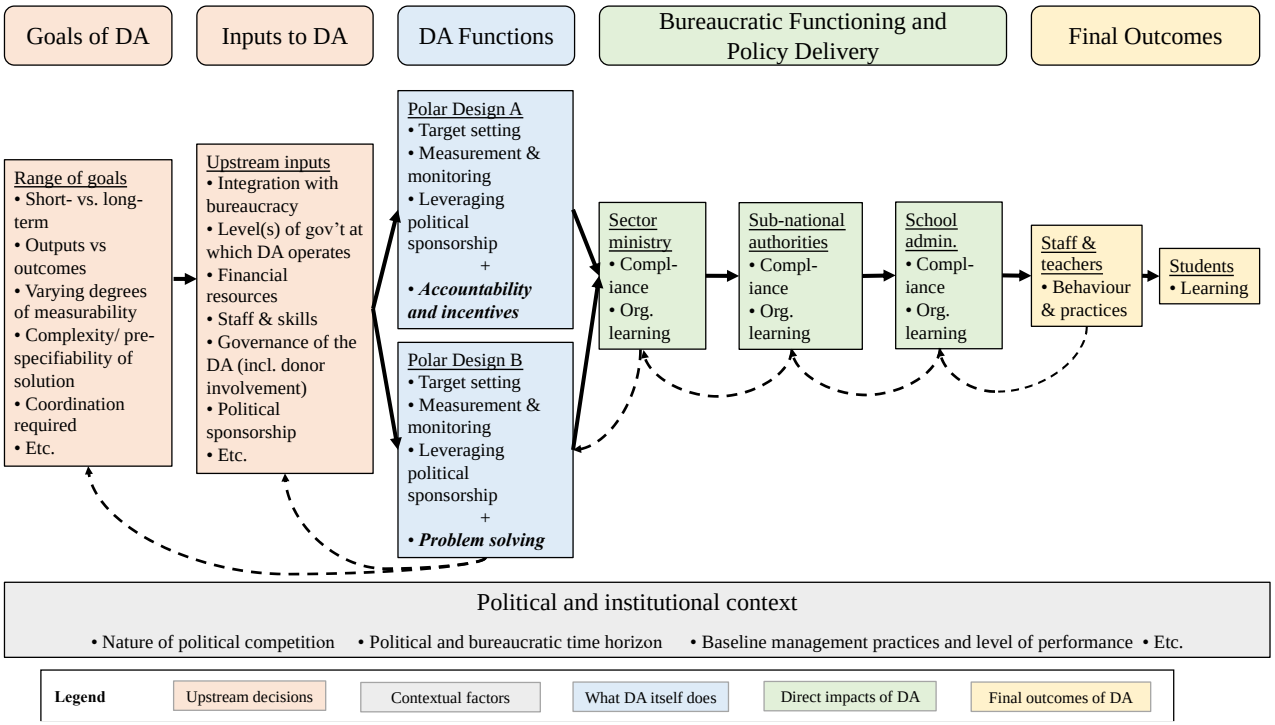
4 This does not exclude delivery units that are co-managed by consultants or contracted out to consultants. This is because in both these instances, the government remains the client.

4. A Theory of Change and Typology 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? A clear understanding of this is necessary both in order to clarify what other variables influence DAs 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 factors to each other, rather than making any recommendations about what role DAs should play or how they should be designed.

Figure 1 depicts our high-level theory of change for delivery approaches (DAs). We focus on the case of the education sector, but the framework can be generalized to any service delivery sector. Figure 1 distinguishes among five key sets of variables: 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 ideal types described in greater detail below; the 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; 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; and changes in the final outcomes of interest (yellow) which are themselves results of the better bureaucratic functioning that might result from the DA's operations, but which are not themselves directly affected by the DA. Fifth and finally, these decisions, actions, and outcomes all occur in a political and institutional context (grey) 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 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 in order to undertake its functions. Second, the DA may have the necessary inputs but carry out its functions poorly. 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 in order 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, but in reality understanding causal effects and implementation failures often requires simultaneous analysis across different levels. There are also potential feedback loops embedded into the system either by design or by accident, of which Figure 1 illustrates some particularly important ones. 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). 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.

Section 5 below discusses the DA functions and the conceptual distinction between the two ideal types they can take. 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. Delivery Approach Functions and Ideal Types

5.1 Delivery Approach Functions

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 which delivery approaches can draw on to achieve their objectives.5 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 (World Bank, 2004; Laffont and Martimont, 2002; Hood, 1991; Locke & Latham, 1960). 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.

5 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.

1) **Target setting and prioritization**: The establishment of a set of key priorities and objectives, measurable indicators to characterize progress against these objectives, and benchmark levels of performance to be achieved in a specified time period.
2) **Measurement and monitoring**: The establishment and execution of mechanisms to collect and report information about 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 is 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 meeting 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 (1) with leveraging political sponsorship (3) to increase its salience, establish higher-frequency measurement of performance (2), and routinise performance review in the presence of the sector minister with naming-and-shaming of good and bad performers (4). 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.

5.2 Two Ideal Type Delivery Approach Designs

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 broad styles of delivery approach: one which relies on top-down incentives and accountability as its key function ('Type A'), and one which relies on horizontal and bottom-up convening, collaboration, and problem-solving as its key function ('Type B'). These key levers differentiate these two different types of approaches, which otherwise share a number of common functions (target setting, measurement and monitoring, leveraging political sponsorship).

The underlying theory behind Type A delivery approaches 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. This style of delivery approach proposes the use of accountability and incentives to mitigate this agency problem and hence to improve policy implementation.

In their conceptually pure form, Type A delivery approaches utilize 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 in order 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.

An empirical example of a delivery approach that relied primarily (though not exclusively) on a Type A approach is the original UK PMDU, which combined high-profile target setting with a top-down, data-driven 'targets and terror' accountability regime (Bevan and Hood 2006). While this delivery unit was also trying to catalyze innovation, innovations were seen more as an intermediate outcome of increased effort – rather than something which the delivery unit itself sought to facilitate or routinise (Richards & Smith, 2006). Although part of the intention behind the establishment of the PMDU was to promote more ownership of delivery and discretion among frontline bureaucrats, the tools and routines the unit leveraged initially concentrated the oversight and responsibility for delivery at the Centre with monitoring against specific targets and deliverables, although this approach shifted somewhat over time (Richards & Smith, 2006; World Bank, 2010a, 2020b; CPI, 2016a).

The underlying theory behind Type B delivery approaches 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. If Type A approaches presume that principal-agent problems (lack of incentives) are the main challenge for policy implementation, Type B approaches presume instead that the main challenges are that 1) disempowered agents are unable to use discretion and flexibility effectively, and 2) there is poor coordination among agents, such that data, evidence, and feedback loops are not incorporated into the decision-making processes both horizontally and vertically in the delivery chain.

Whereas the intellectual origins of Type A approaches are in the more hierarchical Weberian approach to bureaucracies and in incentive-driven New Public Management theories, the intellectual origins for Type B delivery approaches can be found in the development, governance and public administration literatures that emphasize the importance of frontline discretion, innovation, and coordination (Pressman and Wildavsky 1973; Lipsky 1980; Milward 1982; Andrews et al 2013; Andrews 2015; Korten 1980; Argyis and Schon, 1997). Within the education sector, this perspective is grounded in theories of distributed leadership (Coburn, 2016; Leithwood, 2011), organizational learning (Coburn & Honig, 2008; Honig, 2008), and improvement science (Bryk et al., 2010; Huber, 1991). These frameworks draw from a long tradition in the learning sciences and study of human cognition and motivation, which 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 and 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, and Levy 2018).

Like Type A approaches, Type B delivery approaches utilize target setting (function 1), measurement and monitoring (2), and political sponsorship (3) as important tools for improving policy implementation. However, they seek to use these tools not in service of accountability regimes for doling out rewards and sanctions, but as a way to enhance the problem-solving and collaboration routines (function 5) that they seek to institutionalise. The intention is to enable iterative learning and adaptation through collaboration, problem solving, and/or development of professional judgment - processes which change not only 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 actually happening, and what is preventing improved performance, and political sponsorship is used to support prioritization of effort and

common understandings of goals rather than in support of the promise (and threat) of performance-linked incentives.

Malaysia's PEMANDU is an example of a delivery approach that relies more on a Type B approach. It shares with the early phase of the UK's PMDU the establishment of a small number of high-profile, priority targets set out with strong political backing, but its main tool for sparking change is the convening of six-week-long delivery labs that bring together key stakeholders from different sectors and organizations who are collectively responsible (and arguably necessary) for achieving these targets (World Bank, 2017). While normal bureaucratic accountability mechanisms still apply to public servants and organizations that participate in these labs, of course, the collective nature of the target and actions (as well as the involvement of non-governmental stakeholders) make it less feasible for these accountability mechanisms themselves to be the main drivers of change. These delivery labs are followed up with a series of 'Open Days' where implementation plans, milestones, and key targets—presented in Roadmaps—are communicated to the general public and revised based on collected feedback (PEMANDU Associates, n.d.; World Bank, 2017). As beneficiaries of public services, citizens' knowledge about the needs and circumstances of their community is considered valuable information to share upstream with bureaucratic agents and political leaders to inform policies and their implementation (World Bank, 2017; CPI, 2016b).

These ideal types represent hypothetical conceptually pure versions of the underlying concept, but should not be understood as 'ideal' in the normative sense that policymakers should aim to actually operationalize them in their pure form. In practice, almost all actually existing delivery approaches involve some combination of these two types, although most are weighted more towards one type. Of course, the choice of a particular type of delivery approach is a hypothesis on the part of the designers about the nature of the underlying problem it is trying to solve, as well as about the relative effectiveness of different management tools for solving it. In this sense, whether Type A or Type B delivery approaches (or the near-infinite range of potential hybrid approaches) 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.

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).6 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

6 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.

mediating variables of management and policy implementation. This section briefly discusses each set of variables and its 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 Delivery Approach Goals

DAs are set up to achieve a range of goals which 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 in order 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 ideal types.

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 (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 typically 'fidelity' problems, where the objective is to deliver a specific intervention with high fidelity. Goals whose solutions are difficult to pre-specify are typically '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 Delivery Approach 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 three types of (non-)integration:
  - **Existing bureaucracy**: Using existing structures within the bureaucracy to execute the key functions of the delivery approach.
  - **Hybrid or re-organization of bureaucracy:** A hybrid of existing and new bureaucratic structures, or a re-organizing or shuffling of pre-existing bureaucratic structures into a novel configuration.
  - **New unit:** Creation of a new unit that exists outside or in parallel to existing structures of bureaucratic hierarchy and resource allocation.

➔ **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:

- **Centre-of-government:** The top-most level of political and bureaucratic authority, such as a president or prime minister's office.
- **Ministerial:** A national-level sector ministry or other national-level agencies, such as an education ministry.
- **Sub-national:** Any government entity that exists below the national level. This could be states, provinces, regions, districts, or municipalities.
- **School:** 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:

- **Existing staff** operating within normal civil service roles and regulations;
- **New/external staff** such as consultants, technical assistants, or other external hires working outside the mainstream civil service;
- **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 which 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 government to effectively reward or sanction public personnel, and entails clientelistic 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 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, Chijoke, & Barber, 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.

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 which 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, Lemos, & Scur, 2019) to nuanced qualitative data from interviews with key bureaucrats to the coding of administrative data on task or output completion (e.g. Rasul, Rogger, & Williams, 2020). A detailed approach to the study of changes in bureaucratic functioning and capacity will be the focus for a companion paper to this one which will lay out methodological approaches that will be used for the study.

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 back to the distinction between the two ideal types 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, Kain, O'Brien & Rivkin, 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 which succeeds in improving policy implementation but fails to improve outcomes because 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 decisionmakers. The extent to which they try and succeed to do this is an interesting question for empirical study, and might be expected to be more prevalent in 'Type B' approaches (as illustrated by the dotted feedback loops in Figure 1).

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 (grey x blue → green, or grey 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 → 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 Types A and B) 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 ideal types are associated with different types of effects on downstream bureaucratic functioning variables – in particular, whether Type A approaches are relatively more likely to achieve improvements in compliance-oriented functioning and Type B 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?
- Are accountability-intensive (Type A) approaches or problem solving-intensive (Type B) approaches more successful in improving the functioning of the education sector bureaucracy?

- ➔ Do accountability-intensive (Type A) approaches achieve relatively greater improvements in compliance-related functioning, and problem solving-intensive (Type B) 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 have an effect on 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 Type A DAs 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 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 whether Type A or B is relatively more effective 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 (grey x blue → green, or grey x orange → blue → green). This captures the idea that some DA functions or ideal types 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 Type A and Type B approaches?
- ➔ 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.

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 it is our 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. and Thomas, P. (2016). Improving Government Effectiveness Across the World, Institute for Government. Retrieved from: https://www.instituteforgovernment.org.uk/publications/improving-government-effectiveness-across-world

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

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

Andrews, M. (2015). Doing Complex Reform Through PDIA: Judicial Sector Change in Mozambique. Public Administration and Development, 35:4, 288-300.

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. Retrieved from: https://www.newyorker.com/magazine/2014/07/21/wrong-answer.

Baars, Sam, et al. (2014). Lessons from London Schools: Investigating the Success. Centre woofor London and CfBT Education Trust.

Barber, M. (2009). How Government, Professions and Citizens Combine to Drive Successful Educational Change. In Second International Handbook of Educational Change, edited by Andy Hargreaves, Ann Lieberman, Michael Fullan, and David Hopkins. Springer International Handbooks of Education. Springer Netherlands. Retrieved from: https://doi.org/10.1007/978-90-481-2660-6.

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

Bevan, G., and 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.

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.

Buerkli, D. (2015). Welcome to Mindlab: Designer of Delivery and Innovation in Denmark.

Centre for Public Impact (CPI). (2016a). The Prime Minister's Delivery Unit (PMDU) in the UK. Boston Consulting Group. Retrieved from: https://www.centreforpublicimpact.org/case-study/prime-ministers-delivery-unit-uk/. Last accessed June 2nd 2020.

Centre for Public Impact (CPI). (2016b). Malaysia's Performance Management and Delivery Unit. Boston Consulting Group. Web. https://www.centreforpublicimpact.org/case-study/performance-management-delivery-unit-kuala-lampur/. June 2nd 2020.

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, 122:3, 465-475.

Daly, E., Riese, J. and Singham, S. (2015). The Delivery Challenge: A Systematic Approach to Achieving Breakthrough Impact. McKinsey and Company. Retrieved from: https://www.mckinsey.com/industries/public-sector/our-insights/the-delivery-challenge-a-systematic-approach-to-achieving-breakthrough-impact.

Daly, E. and Singham, S. (2012). Delivery 2.0: The New Challenge for Governments.

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

Dumas, V, Mariano Lafuente and S Parrado. (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. Retrieved from: https://publications.iadb.org/en/strengthening-center-government-results-chile-experience-ministry-presidency-and-its-presidents. Last accessed June 2nd 2020.

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

Freeguard, G and Jen Gold. (2015). Data-driven delivery: lessons from the O'Malley Administration of Maryland. Institute for Government. Retrieved from: https://www.instituteforgovernment.org.uk/publications/data-driven-delivery-lessons-o%E2%80%99malley-administration-maryland.

Gold, Jen. (2014). International Delivery: Centres of Government and the Drive for Better Policy Implementation. Institute for Government. Retrieved from: https://www.instituteforgovernment.org.uk/sites/default/files/publications/International%20Delivery%20report.pdf. Last accessed June 2nd 2020.

Gold, Jennifer. 2017. "Tracking delivery: global trends and warning signs in delivery units. Institute for Government." Institute of Government. Web. https://www.instituteforgovernment.org.uk/sites/default/files/publications/Global%20Delivery%20report.pdf. Last accessed June 2nd 2020.

Glewwe, P., and Muralidharan, K. (2015). Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications. RISE. Retrieved from: https://www.riseprogramme.org/sites/www.riseprogramme.org/files/publications/RISE_WP-001_Glewwe_Muralidharan.pdf. Last accessed June 2nd 2020.

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, 24:3 133-50.

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, 627–64. Retrieved from: https://doi.org/10.1086/589317.

Honig, M.I., and C. Coburn. (2008). Evidence-Based Decision Making in School District Central Offices: Toward a Policy and Research Agenda. Educational Policy 22:4, 578–608. Retrieved from: https://doi.org/10.1177/0895904807307067.

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., and 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., and Moody, C. (2016). "What is a Delivery Unit?" Deloitte Delivery Results Series. Web. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/us-fed-what-is-a-delivery-unit.pdf. Last accessed June 2nd 2020.

Korten, D. (1980). Community Organization and Rural Development: A Learning Process Approach. Public Administration Review, 40:5, 480-511

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

Laffont, J., and Martimort, D. (2002). The Theory of Incentives: The Principal-agent Model. Princeton, N.J.

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge university press.

Leaver, C., Lemos, R., and Scur, D. (2019). Measuring and Explaining Management in Schools: New Approaches Using Public Data. RISE. Retrieved from: https://assets.publishing.service.gov.uk/media/5ed0e8f5d3bf7f45fcf2c5b2/RISE_WP-033_Leaver_Lemos_Scur.pdf. Last accessed June 4th 2020.

Leithwood, K., & Seashore-Louis, K. (2011). Linking Leadership to Student Learning. John Wiley & Sons.

Levy, B., and Kpundeh, S. J. (2004) Building State Capacity in Africa: New Approaches, Emerging Lessons. World Bank Publications.

Levy, B., Cameron, R. Hoadley, U. and Naidoo, V. eds. (2018). The Politics and Governance of Basic Education: A Tale of Two South African Provinces. First edition. Oxford, United Kingdom: Oxford University Press. Retrieved from: http://fdslive.oup.com/www.oup.com/academic/pdf/openaccess/9780198824053.pdf.

Lindquist, E.A. and Wanna, J., (2010). 1. Delivering policy reform: making it happen, making it stick. Delivering policy reform .

Milward, Brint. (1982). Interorganizational Policy Systems and Research on Public Organizations. Administration and Society 13:4, 457-478.

Mourshed, M., Chijoke, C. and Barber, M. (2007). How the School's Most Improved School Systems Keep Getting Better. Mckinsey. Retrieved from: https://www.mckinsey.com/~/media/McKinsey/Industries/Social%20Sector/Our%20Insights/How%20the%20worlds%20most%20improved%20school%20systems%20keep%20getting%20better/How_the_worlds_most_improved_school_systems_keep_getting_better.ashx. Lasr accessed June 2nd 2020.

Nordstrum, L. E., LeMahieu, Paul G., Dodd, K. (2017). Deliverology in Kentucky. Quality Assurance in Education, 25:1, pp.43-5.

Opper, I. M. (2019). Teachers Matter: Understanding Teachers' Impact on Student Achievement. Santa Monica, CA: RAND Corporation. Retrieved from: https://www.rand.org/pubs/research_reports/RR4312.html. Last accessed June 2nd 2020.

PEMANDU. (n.d.). Big Fast Results (BFR) Methodology: 8 Steps of Transformation. PEMANDU Associates. Web. https://www.pemandu.org/8-steps-of-transformation/. Last accessed June 2nd, 2020.

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

Pritchett, L. (2013). The rebirth of education: Schooling ain't learning. CGD Books
Rasul, I., Rogger, D., & Williams, M. (2020). Management, Organizational Performance, and Task Clarity: Evidence from Ghana's Civil Service. Mimeo, June.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. American economic review, 94:2, 247-252.

Richards, D., and 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. Retrieved from: DOI: 10.1080/13876980600971151. Last accessed June 2nd 2020.

Sanders, W. L., & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student academic achievement.

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

Scharff, M. (2013). Translating Vision into Action: Indonesia's Delivery Unit, 2009-2012. Innovations for Successful Societies. Princeton, NJ: Princeton University .

Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T., ... & Eyers, J. (2015). 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, London.

Shostak, R., J. Watkins, A. Bellver and I. John-Abraham. (2014). When Might the Introduction of a Delivery Unit be the Right Intervention? World Bank Governance and Public Sector Management Practice Notes.

Todd, R. and Waistell, D. (forthcoming). Overview of the Delivery Approach. Cambridge Education report for the Bill and Melinda Gates Foundation.

Vygotsky, L. S. (1978). Socio-cultural theory. Mind in society.

Woolcock, M., & Prichett, L. (2002). Solutions When the Solution is the Problem. Arraying the Disarray in Development, Washington, DC.

World Bank. (2004). World Development Report: World Bank World Development Report 2004. Development and Change 36:6, 1263-266. Retrieved from: https://openknowledge.worldbank.org/handle/10986/5986. Last accessed June 4th 2020.

World Bank. (2010a). Improving Performance : Foundations of Systemic Performance. Global Expert Team (GET) Brief. World Bank, Washington, DC. Retrieved from: http://documents.worldbank.org/curated/en/923311468337217269/pdf/600920BRI0GET010BOX358310B01PUBLIC1.pdf. Last accessed June 3rd 2020.

World Bank. (2010b). Ghana's New Approach to Public Sector Reform: Focusing on Delivery. 70156. Washington, D.C.: World Bank. Retrieved from: https://openknowledge.worldbank.org/bitstream/handle/10986/12811/701560ESW0P1230Delivery0Model0Note.pdf?sequence=1&isAllowed=y.

World Bank. (2017). Driving performance from the center : Malaysia's experience with PEMANDU (English). Washington, D.C. : World Bank Group. Retrieved from: http://documents.worldbank.org/curated/en/318041492513503891/Driving-performance-from-the-center-Malaysia-s-experience-with-PEMANDU. Last accessed June 2nd 2020.

World Bank. (2018). World Development Report: Learning to Realize Education's Promise. World Bank Group. Retrieved from: https://www.worldbank.org/en/publication/wdr2018. Last accessed June 4th 2020.