

Targeting Results, Diagnosing the Means: Innovative approaches for improving public sector delivery

Nick Manning, Joanna Watkins¹

98815

In some ways, public policy is the easier part of development. Much is known about what, ideally, is to be achieved in health, education, infrastructure etc. When and how the public sector can be helped to implement those agreed policies is less clear. The “binding constraint” on development is the capacity of the public sector to deliver effectively.

In seeking to improve delivery capacity, the practitioner and academic literature on public sector reform in developing countries is increasingly focused on results: it argues that there is little point in judging the public service by the degree to which it adheres to some notion of procedural or institutional best practice – the public sector is as good as what it achieves, not what it looks like.

This note sets out approaches to reform which start with identifying the shortcomings in results and which then look for pragmatic solutions that fit the particular context: no best practice, fewer universal recommendations for institutional design. The relative merits of this type of approach have not been empirically tested, but they are nonetheless intuitively reasonable and offer an alternative to other models of institutional reform which have not had great success.

This note sets out three recent examples of this new approach. The first, “Deliverology,” focuses on defining ambitions and planning for delivery improvements with an improvement trajectory mapped out (Barber et al. 2011b; Barber 2008; Barber et al. 2011a). Three critical elements of this approach include the formation of a delivery unit, data collection for target setting, and the establishment of management routines. The Problem Driven Iterative Adaptation (PDIA) (Andrews et al. 2012) aims to solve problems as understood by local actors with experimentation and the iterative feedback of lessons into new solutions. The World Bank’s Public Sector Management Approach (World Bank 2012c) has its three principles of solutions based on rigorous diagnostics, agile implementation and learning as we go. These approaches all emphasize the initial focus on results and the “whatever it takes” approach to the underpinning institutional reforms.

GET Notes – Recently Asked Questions Series intends to capture the knowledge and advice from individual engagements of the World Bank’s Global Expert Team on Public Sector Performance (PSP GET). The views expressed in the notes are those of the authors and do not necessarily reflect those of the World Bank. For more information about the PSP GET, contact the GET team leader Bill Dorotinsky (wdorotinsky@worldbank.org) or visit: <http://www.worldbank.org/pspget>

¹ This note was prepared by Nick Manning with Joanna Watkins drawing on extensive assistance from: Ana Bellver, Theo Thomas, Frederico Gil Sander, Willy McCourt (Delivery Units); Vivek Srivastava, Jurgen Blum (indicators); Francesca Recanatini, Verena Fritz, Jurgen Blum, Robert Beschel (political economy); and Tony Verheijen and Willy McCourt (World Bank strategy).

This note argues that these results-based approaches are a welcome breath of fresh air in a difficult domain. They are clearly in tune with the current results focus of the international development community and they address many of the challenges recognized by practitioners in previous approaches. However we still have remarkably little hard evidence on which to base a robust assessment of the effectiveness of this type of intervention. It looks and feels encouraging – but in reality, all we know right now is that it is a significant departure from plan A. We do not know that, as a plan B, it is a success.

The note concludes with three questions:

Question 1: How could we build research and evaluation into this broad family of results-based reform approaches?

Question 2: Within any overall evaluation of these approaches, how can some key tactical questions be examined:

1. Are “Delivery Units” or equivalent important for “anchoring” reforms?
2. Does defining the performance problem at the local level, with minimal top-down involvement, result in a focus on small scale changes?
3. What is the impact of a reliance on quantitative performance targets?
4. Must these approaches always be completely agnostic about “upstream” public sector systems and design?

Question 3: How can donors be motivated to experiment with new modalities in an area where there is a long tradition of focusing on the best practices to be transferred rather than the delivery problems that are to be solved?

I. A New Approach to Public Sector Reform?

The costs of moving from project or policy design to effective public sector implementation are surprisingly varied. For example, Nigeria spends about four times as much for health per capita as Ethiopia – but more under-5s die every year in Nigeria than in Ethiopia.ⁱ Why? Ghana and Benin have similar per-capita income levels, but in Ghana, 15 to 24-year olds literacy rates are about 50 percent higher than in Benin.ⁱⁱ Why? Why is total health spending in the US nearly twice as high as in Sweden, yet infant mortality in Sweden is less than half as high as in the US?ⁱⁱⁱ Why does a typical unit of spending in developing countries translate into just half of that value when invested in physical capital assets (Gupta and Verhoeven 2001)?

The key variable here is not policy in the sense that different outcomes were being sought concerning health care, educational attainment or productive infrastructure. The question is the ability of the public sector to ensure delivery - the degree to which the public sector is able to implement agreed policies.

In seeking to improve delivery capacity, the practitioner and academic literature on public sector reform in developing countries is increasingly focused on results. The weight of reform recommendations is shifting to the view that there is little point in judging the public service by the degree to which it adheres to some notion of procedural or institutional best practice (form); instead, we should start by looking at shortcomings in delivering results (function) and find the

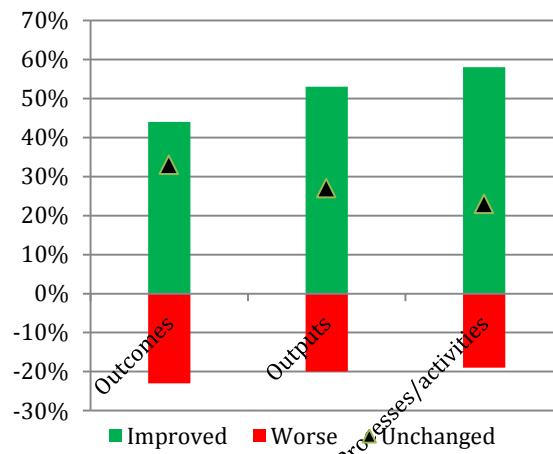
arrangements that best fix the problem. So, whatever the output – whether it is services such as health and education, the management of infrastructure and other public investments, the regulation of social and economic behavior, the setting of sector policy objectives, ensuring fiscal and institutional sustainability – the public sector is as good as what it achieves. Therefore, starting with the identification of shortcomings in results and then finding the fix that fits is increasingly seen as the way to go.^{iv} For shorthand, these approaches will be referred to subsequently in this note as “targeting results, diagnosing the means” approaches (TRDM).

Replacing the question of what the public sector should look like (what is the ‘best practice’ for this organization?) with consideration of what it is not achieving and what fix might fit is a significant departure from previous technical assistance recommendations. In caricature, that previous approach was based on a view that particular institutional forms were universally the right way to go – and that they could always be introduced regardless of the context.^v Whether this TRDM approach is really being embedded into donor advice, and whether the earlier approach really was as strongly best practice oriented as is often maintained, is hard to assess. But the *rhetoric* surrounding technical assistance for public sector reforms in developing countries has certainly changed.

Like most points of view in public sector reform, the arguments in favor of this new TRDM orientation have relatively little hard evidence supporting them^{vi} but they are intuitively reasonable. They also rest on some robust empirical findings that starting with best practice institutional reform and building from there has not been a great success, even in relatively robust governance environments (see **Error! Reference source not found.** on the mixed evidence of New Public Management Reforms in the European Union). “The fact that the “development community” is five decades into supporting the building of state capability and that there has been so little progress in so many places (obvious spectacular successes like South Korea notwithstanding) suggests the generic “theory of change” on which development initiatives for building state capability are based is deeply flawed.” (Andrews et al. 2012, 2)

These new TRDM approaches address three challenges in public sector reform and development, well-known to PSM practitioners. First, it is surprisingly hard to know that decisions about new processes and reformed systems have been implemented in practice. A new civil service law or new budgetary procedures can be proposed and agreed, but implementing a new merit-based promotion policy within the civil service requires changing the hard-to-observe behavior of thousands of public

Figure 1: The success and failure of “New Public Management” reforms in the European Union



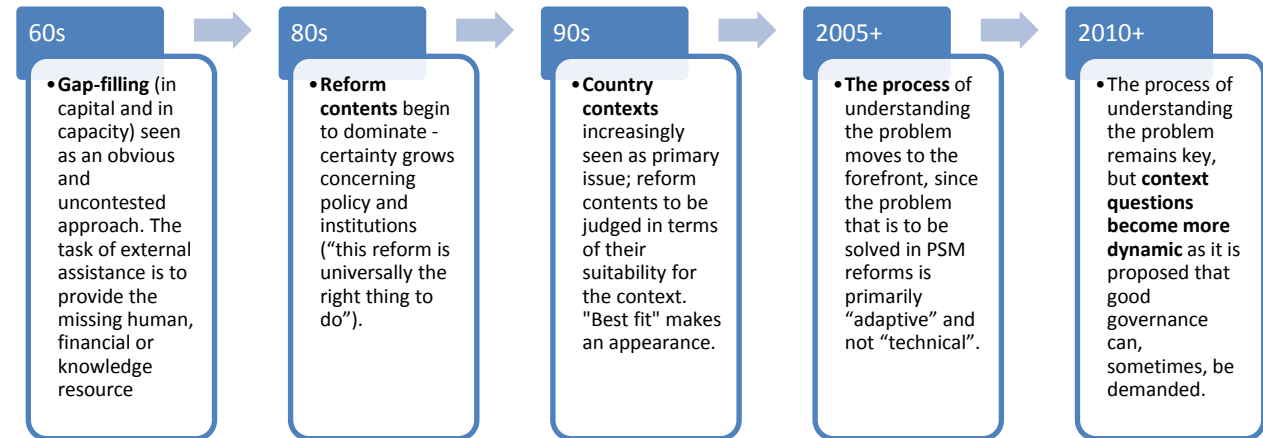
Source: (Pollitt and Dan 2011)

servants, many of whom can continue patterns of patronage while claiming to have introduced the policy wholeheartedly. If the introduction of a managerial “best practice” is the objective, then it is much harder to know that this has actually happened than it is to know whether children are now being vaccinated. This problem of unobserved behaviors is exacerbated by the political stakes involved in reform. Many politicians promise improved results, while very few seek election on an administrative reform platform. Making changes in the way that money and people are managed within the public sector can be troubling for many interest groups and so there are many political temptations to collude with Potemkin Village-like managerial reforms that have little real significance in practice.^{vii} The TRDM approach offers, at least in principle, a way of minimizing the risks that actors deep within the public sector can “fool” the reformers by just going through the motions of reform without really changing their behavior since the reforms are verified in the driving question about performance improvements.

Second, the results chain between upstream decisions and downstream service delivery or other results is long and complicated. Even if managerial reforms are implemented in practice, there is no performance gain if other weak links in the chain represent more fundamental obstacles. For example, introducing a school-based management regime may improve management of resources, but will have little impact on learning outcomes if poor quality teaching staff is the binding constraint. Managerial reforms that have only a marginal impact on the problem at hand are largely wasted effort. A TRDM approach reduces the risk that the real constraints are being missed and reform effort wasted by asking the question “what is the smallest set of changes that have to be implemented in order to ensure that this result is achieved?”

Finally, since TRDM approaches are essentially focusing on the achievement or result and are agnostic about the reforms that will lead there, they allow for the possibility of adaptation in the approach during the reform. How the system will be redesigned is not locked in place at the time of planning the reform – minimizing the risk that there will be loss of face for the reformers if they have to change their original assumptions. This structured agnosticism is consistent with the widespread recognition in development in general, and in public sector reform in particular, that there is less known than was previously thought – and that there is a strong need to adapt approaches during the reforms that they are supporting (Figure 2).

Figure 2: Changing development ideas concerning external influences on Public Sector Management^{viii}



Source: (Blum et al. 2012)

This line of thinking can be seen within the arguments for Cash on Delivery Aid (Birdsall and Savedoff 2010) and the World Bank's Program for Results instrument (World Bank 2012b), although as noted below, the politics of change within donor organizations are not always straightforward.

A. TRDM approaches seem to be taking off

Recent high profile proponents of multi-sector TRDM approaches include:

- Deliverology (Barber et al. 2011b; Barber 2008; Barber et al. 2011a)
- Problem Driven Iterative Adaptation (PDIA) (Andrews et al. 2012; Andrews 2013b)
- The World Bank approach to Public Sector Management 2011-2020 (Blum et al. 2012; World Bank 2012c)

These approaches are at very different levels of generality. "Deliverology" describes a set of interventions with some real world examples of their consequences (government-wide in Malaysia, Indonesia, Sierra Leone and others; education sector in Punjab etc.). Both the PDIA and the World Bank approaches draw on elements that can be seen in current projects, but they are essentially normative proposals rather than strategies with a track record that can be evaluated. Reform approaches in Sierra Leone (Srivastava and Larizza 2012) and in Punjab (World Bank 2012a) have been cited as examples of the World Bank approach in practice. PDIA approaches are referred to in recent reforms of Ministries of Finance in the Caribbean (Brown and Smith 2013) and in reform approaches in Mozambique (Andrews 2013a) and in Burundi (Andrews 2013c).

Box 1 sets out their key positions. These approaches all emphasize the initial focus on results and the "whatever it takes" approach to the underpinning institutional reform. However, behind that headline there are as many differences as similarities, as seen in Table 1.

Box 1: TRDM “manifestos”

“**Deliverology**” as a set of questions (Barber 2008, 73)...

- What are you trying to do?
- How are you trying to do it?
- How do you know you are succeeding?
- If you’re not succeeding, how will you change things?
- How can we help you?

...developed into a systemic approach (Barber et al. 2011b)

- Develop a foundation for delivery - by defining your aspirations, reviewing where you are now, defining who is going to help in achieving these goals
- Understand the delivery challenge – by evaluating past and present performance and understanding what drives and what constrains performance within the system
- Plan for delivery improvements - determine the strategy and targets, and set the improvement trajectory that you will get onto
- Drive delivery – with routines to drive and monitor performance, solving problems promptly as you go, with an emphasis on building momentum

The **Problem-Driven Iterative Adaptation** (PDIA) approach is based on core principles consistent with a wide range of implementation options rather than a specific single program or approach. This approach aims to solve particular problems in particular local contexts via:

- i. the creation of an ‘authorizing environment’ for decision-making that encourages experimentation and ‘positive deviance’, which gives rise to
- ii. active, ongoing and experiential (and experimental) learning and the iterative feedback of lessons into new solutions, doing so by
- iii. engaging broad sets of agents to ensure that reforms are viable, legitimate and relevant—that is, are politically supportable and practically implementable.

(Andrews et al. 2012, 8)

The **World Bank’s Public Sector Management Approach** (World Bank 2012c) seeks to bridge the gap between what we know about PSM reform and how we do it:

- Principle 1: Design solutions based on rigorous diagnostics

Start with a degree of agnosticism on what works. Good diagnostics focus on solving a performance problem; they engage stakeholders in the problem-solving process and start with the hypotheses that the status quo of public institutions represents a *functional* equilibrium given stakeholder incentives, even if this equilibrium may be very dysfunctional for ends such as service delivery.

- Principle 2: Implement with agility

The traditional emphasis on well-defined but rigid project design is challenged by the growing evidence that implementation processes matter crucially for results – in particular in building public sector institutions. If experimentation and learning-by-doing are increasingly seen to be key to success, the traditional distinction between “design” and “implementation” in reform projects can be a constraint.

- Principle 3: Learn as we go

Practitioner’s experience is of course an invaluable source of knowledge for PSM reform design – many senior administrators and advisors can sense that a reform is implausibly ambitious or excessively modest. However, by itself, tacit knowledge held by experts is insufficient – there is a need for constant empirical testing of what works in reform.

(Blum et al. 2012)

Table 1: Differences in TRDM approaches

	Relation to best practice models	TRDM Approaches		
		Problem Driven Iterative Adaptation	Deliverology	World Bank’s Public Sector Management Approach
How is the reform process anchored?	Project-specific, but with an emphasis on commitment from senior political levels	Project/problem-specific	Often involves a Delivery Unit or equivalent, close to head of government	Project-specific, but with an emphasis on commitment from senior political levels
How is the problem defined?	Deviation from ‘best practice’ is the problem	Local definition of the performance problem – central/top-down problem definition is minimized	The problem is defined through a Priority Review. The head of government (or top management level) is involved in the definition of delivery failure.	Bank and counterpart team to undertake diagnostic work on functional failure, likely to include central/top-down definitions of the problem
How is the action plan determined?	Engineering model of implementation plan – Gantt charts and clear specification of components and phasing	“Muddling through” emphasized – specifications kept to the minimum combined with encouragement for innovative ideas to emerge during implementation	Formal evaluation of past and present performance and feasibility of improvements – trajectory defined so that progress improvements can be kept on track	Open discussion with stakeholders combined with more Bank-centric prospective political economy analysis (how will the actors react to these changes?); significant use of empirical data to assess feasible levels of ambition
How much does implementation rely on quantitative targets?	Indicators of compliance with the implementation plan are key	Targets likely to be softer and include qualitative considerations	Very significant use of quantitative targets	Moderate use of quantitative targets for measuring results and for measuring the strength of public management systems (assumed to be important for sustainability)
How flexible is implementation of the plan?	Limited flexibility – the need to change implementation plan seen as failure of original project design	Implementation changes welcomed Tight feedback loops to see if the changes are addressing the problem and rapid changes if not	Moderate flexibility – clear timebound plan, with targets for delivery improvements set out Plan updates are feasible but not to be taken lightly	Moderate flexibility – project redesign as part of implementation highlighted, drawing on stakeholder ideas and experimentation Results-based lending and possible use of ICT for rapid feedback highlighted
What locks in the performance improvements?	The problem identification is assumed to include an assessment of what is needed to maintain the delivery gains	The depth of ownership of the reforms by the local actors, and the degree to which the reforms assist in addressing problems which they themselves identified, is assumed to entrench the reforms sustainably	Key drivers of sustained performance improvements are: trained staff; institutionalized performance data collection regime; and enthusiasm/momentum deriving from early successes	Delivery improvements in the sectors assumed to be sustained by foundational improvements in core public management systems (HR, PFM, etc.)
How are results scaled up?	Top-down – formal project evaluations disseminated through usual organizational channels	Peers and communities of practitioners involved in the process of problem definition and implementation – learning is organic	Centrally-driven push on delivery improvements maintained – lessons incorporated in routines recommended by the delivery unit.	Dual emphasis on: extraction of tacit knowledge via practitioner networks; and empirical testing of what works in reform with an emphasis on cross-country, cross-sector comparative data and case studies

Adapted from (Andrews et al., 2012; Barber et al., 2011b; Barber, 2008; Barber et al., 2011a; Blum et al., 2012; World Bank, 2012c)

B. What should we take into account in assessing if these approaches work?

These TRDM approaches are intended to inform reform strategies which will drive measurable improvements – they are not models for organizing the public sector. The key questions are when and under what circumstances do they contribute to an improvement in delivery – not what the original levels of performance were.

Realistically, we are not going to see a large dataset of well-measured TRDM interventions which can be contrasted with other, more “best practice” approaches any time soon. An initial understanding of the significance of the approach will more likely be obtained from case studies which analyze the impact of these approaches on different problems within different country contexts.

Country and agency contexts differ in terms of the feasibility of service delivery improvements – the likelihood that any intervention will make any difference in the short term. The case studies will need to show how TRDM and best practice approaches have played out in contexts where these key dimensions of difference within the public sector are well understood.

Box 2: Core public management systems

Budgetary and financial management system

Subsystems

1. Planning and budgeting
2. Financial management
3. Accounting, fiscal reporting and audit

Procurement system

Subsystems

1. Quality management in legislations and regulations
2. Capacity development
3. Operations and market practices
4. Transparency

Revenue mobilization system

Subsystems

1. Tax policy
2. Tax administration

Public administration system

Subsystems

1. Management of operations within the core administration
2. Quality management in policy and regulatory management
3. Coordination of the public sector HRM regime outside the core administration

“Public information” systems

Subsystems

1. Access for citizens to information
2. Monitoring and evaluation (M&E) framework for sector ministries

Source: (PRMPS 2012b; PRMPS 2012a).

Arguably, there are four key dimensions of difference which are likely to impact the predisposition of the public sector for significant short term improvement.

1. The strength of the core public management systems

Core “public management systems” are, in essence, the key management and oversight responsibilities of the core ministries and agencies at the center of government^{ix} (“upstream”) which have functions that cut across sectors and are broadly seen to matter for “downstream” public sector results and development outcomes. Most would agree that these functions include budgetary and financial management systems, procurement and revenue mobilization systems, and public administration (see Box 2).

2. *The governance environment*

How and whether reforms are implemented depends on the interest of political elites which is inevitably driven by self-interest – remaining in power or in office and self-enrichment. This self-interest is more pressing in weak governance environments with “extractive institutions” (Acemoglu and Robinson 2012) or based around “Limited Access Orders”, where the consensus about rent distributions between elites is unstable (North et al. 2007). In caricature, politicians can choose public sector reform strategies anywhere on a spectrum between two extremes: at one end there is “stewardship” (building for the long term through, for example, merit-based recruitment) and at the other “clientelism” (for example handing out public jobs to friends and supporters). In practice, the motivation of most politicians fall somewhere between these two extremes; for example, many politicians seek to deliver some public goods, while at the same time they may be both unwilling to forgo appointment primarily based on loyalty (rather than competency/merit) or be unable to make certain changes since they do not want to go against the vested interests connected to them. The mix of strategies they pick depends on their motivations, as influenced by political time horizons and survival strategies. Importantly, in a majority of countries, political time horizons and survival strategies are not fully supportive of public sector reforms and improved service delivery (Box 3). Public sector reform efforts have to contend with the constraint of limited time horizons and political calculations, but can seek more opportunities to engage within these in ways that support better outcomes.

Box 3: Political time horizons and survival strategies

Time horizons: Most politicians have limited time horizons; by design where politics is competitive or by default. In some situations, time horizons become very short, when governments frequently come and go due to instable coalitions, coups, or other factors. Very short time horizons make a ‘predatory’ – or self interested - approach to governance more likely. ‘Typical’ political time horizons of 4 to 8 years may still fall short of the time needed to initiate, implement, and reap the results from public sector reforms – affecting their political attractiveness.

Survival strategies: Politicians face choices between delivering benefits to smaller groups of supporters versus the wider public. If economic power, use of violence, and political influence are concentrated in the hands of a narrow elite, politicians will be more likely to target benefits narrowly to these elites through “clientelistic” strategies in order to obtain their support in exchange for public contracts, regulatory privileges, etc. Where sources of de facto power are more broadly distributed, they may have an interest in prioritizing broader public goods (education, public health, etc.). However, even in this latter case, it may be attractive for politicians to try to target rewards – because of a weak political system where the strength of political support is not determined by open debate about policy effectiveness due to the strength of ethnic cleavages (Banerjee and Pande 2009) or confusion about who holds the power (Sacks 2011).

3. *“Last mile” service delivery arrangements*

The sector or service-specific arrangements for delivering a particular service can differ significantly. The principal “last mile service” delivery arrangements are as follows (developed from (Bevan 2012; Le Grand 2007)):

- **Trust & Altruism:** reliance on professional standard-setting and self-regulation (e.g. the traditional dominance of teachers and doctors in the management of health and education services)

- **Hierarchy & “Intelligence”:** the general provision of performance information but with no particular incentives attached to it (e.g. the relatively loose performance-*informed* program budgeting structure in many settings including the Russian Federation)
 - **Hierarchy & Targets:** performance-*driven* budgeting with a requirement to report on performance expectations in budget and on results in entity reports with more or less mechanical consequences (e.g. the No Child Left Behind legislation in the US, UK National Health Service (NHS) reforms)
 - **Choice & Competition:** money follows choice combined with supply-side flexibility (e.g. Charter schools)
 - **Voice and Public Ranking:** naming and shaming (e.g. citizen scorecards in the Philippines)
- See **Error! Reference source not found.** for further details.

4. Availability of skilled human capital

The availability of skilled human resources to staff the public sector varies significantly. Challenges in the supply of skills are evidenced by:

- low educational qualifications within the civil service
- a high number of expatriate advisors or consultants in senior positions
- significant vacancies or vacancies filled with unqualified personnel

In countries already constrained by a limited human resource pool, public sector design factors that impede the ability to attract and retain skilled personnel become particularly problematic. Some of these factors include recruitment policies that are influenced by patronage considerations; compensation policies that do not attract qualified staff; and poor management systems that lack performance appraisals, supervision, or an adequate accountability framework. With considerable caution, diversification of remuneration norms on a limited and carefully monitored basis may be necessary and desirable when technical skills are in short supply.

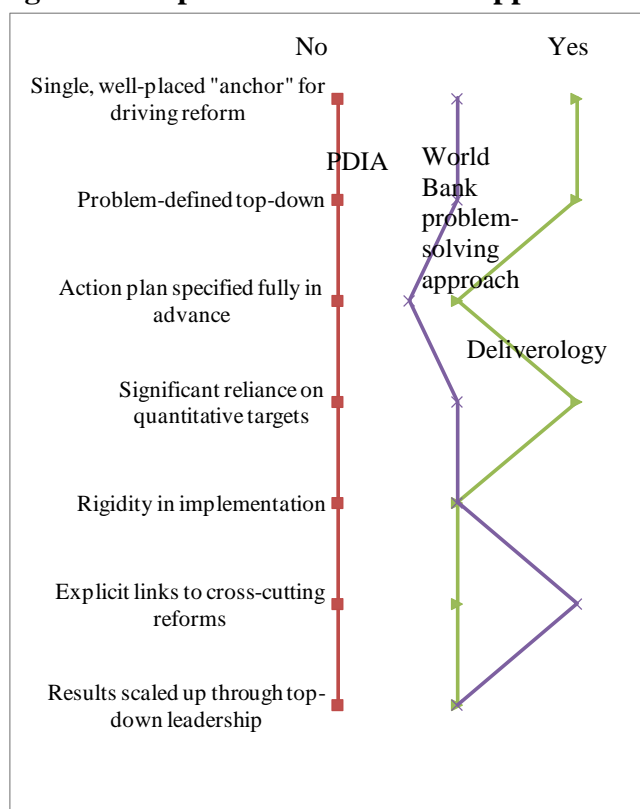
C. What is the significance of the differences within TRDM approaches?

Putting the divergent TRDM approaches in stylized form highlights the differences in emphasis (Figure 3).

The PDIA appears the most radical of these approaches. There are several interesting questions suggested by the areas where the difference between these approaches is the most marked:

1. Are “Delivery Units” or equivalent important for “anchoring” reforms?
2. Does the PDIA approach to defining the performance problem at the local level, with minimal top-down involvement, result in a focus on small scale changes?
3. What is the impact of the reliance on quantitative targets in “deliverology”?
4. Do the links to cross-cutting reforms in the World Bank problem-solving approach embed a partial return to “best practice”?

Figure 3: Emphases within TRDM approaches



In addressing these questions, it should be emphasized that there is a relatively limited evidence base. There is some significant practitioner tacit knowledge – but we are left with much room for implicit assumptions and prior certainties.

1. Are “Delivery Units” or equivalent important for “anchoring” reforms?

Several governments around the world have recently established “delivery units” at the center of government to drive performance improvements to fast track the delivery of public services (see **Error! Reference source not found.**). The Bank has had in depth contact with several of these, including the UK Prime Minister’s Delivery Unit (PMDU), Malaysia’s Performance Management Delivery Unit (PEMANDU), Indonesia’s Delivery Unit (UKP4), Chile’s Presidential Delivery Unit (Unidad Presidencial de Gestión del Cumplimiento), as part of its country dialogue and to facilitate peer learning.

Anecdotally, Bank staff report that Delivery Units are effective to the extent that they maintain a narrow focus and, specifically that they track only a modest set of priority targets^x and *do not*:

- tackle broader civil service reform or budget process issues: while a delivery unit might be well placed to make observations on either of these, their focus is on removing specific bottlenecks; broader public service reforms are important but often take time and require a wider set of stakeholders;
- substitute for the planning or policy functions elsewhere in government – both line ministries and the executive have separate units focusing on these upstream processes, while a delivery unit’s focus is more downstream;
- operate comprehensive monitoring systems— a delivery unit requires a more high-frequency and selective monitoring and reporting framework;
- reopen discussions on the annual budget, as this would likely undermine the main budget process by opening up the possibility of an “end run” around the overall planning and budgeting exercise.

2. *Does defining the performance problem at the local level, with minimal top-down involvement, result in a focus on small scale changes?*

PDIA emphasizes small steps. “In thinking of what (reform)... process should look like, we are reminded of theoretical arguments about how policy and institutional solutions often emerge; as a puzzle, over time, given the accumulation of many individual pieces. Modern versions of such a perspective are commonly called incrementalism or gradualism... The approach holds that groups typically ‘find’ institutional solutions through a series of small, incremental steps...” (Andrews et al. 2012, 13).

This has an air of realism and it is striking that, while reform advocates tend to refer to the need for transformational change in the public sector, reform evaluations tend to praise modest incremental change (World Bank 2003; Independent Evaluation Group 2008). Given the difficulties involved, reform within the public sector probably defaults to the incremental, rather than transformational.

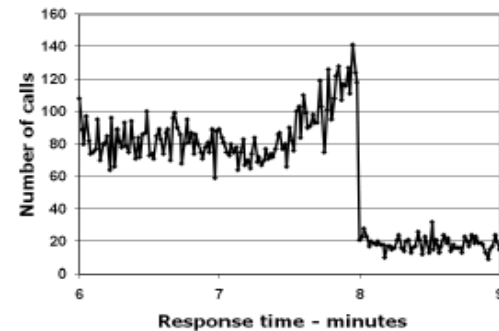
3. *What is the impact of a reliance on quantitative performance targets?*

“Gaming” is universal in any incentive system. There are three main types: output distortion (hitting the target but missing the point/“teaching to the test”); cream-skimming (changing results at the threshold and ignoring the rest); and the ratchet effect (just enough change to trigger the reward, but less than what could be achieved). Many researchers have found empirical evidence of gaming in systems that emphasize incentives for meeting explicit targets. A notable example of this is the effect of a target on ambulance response times that led to many calls being recorded at under eight minutes because staff felt under pressure to record the ‘right’ answer. (See **Error! Reference source not found.** and (Hood 2006; Bevan and Hood 2005; Radin 2006)).

Deliverology emphasizes the establishment of high profile, well-publicized targets – with frequent measurement in order to assess compliance. The focus on what is readily measureable can reinforce output distortions and encourage risk aversion in organizations (Hood 2010). The stated reason for the transformation of the UK PMDU by the incoming coalition government in October 2010 into a unit with a more narrow remit within HM Treasury was the concern that government had become too centralized and focused on top-down compliance targets that were not responsive to users varying needs. This followed widespread concern that resources of front line service providers, teachers and health workers, were being diverted to maintain a cumbersome upward reporting management framework.^{xi}

Figure 4: Frequency distribution of ambulance response times for life-threatening emergency calls in UK

(75% of emergency calls concerning immediately life threatening situations were to be responded to within 8 minutes)



Source: (Bevan and Hamblin 2009)

This problem may be even more acute in countries like Indonesia and Thailand, where the performance framework under the delivery units is less well developed and it can be more difficult to focus on a few key priorities. For example, in Indonesia, there has been general frustration that the delivery unit (UKP4) has not been able to help address many of the underlying weaknesses in the system. For many implementing agencies UKP4 constitutes an additional reporting burden on a system already overwhelmed with largely unused monitoring and evaluation tools, and the focus remains largely on compliance around transactions. As a result, the Government has set up other temporary structures to try to address specific delivery problems, for example: to improve the effectiveness of the national poverty reduction strategy overall oversight and coordination responsibilities have been elevated to a cabinet-level team led by the Vice President (TNP2K); a separate high-level inter-ministerial team has been created to monitor and seek to overcome budget execution problems (TEPPA); while civil service reform is being addressed by a new structure under the President. This burgeoning of institutional forms with corresponding rules, regulations, and processes may undermine the very achievement of results.

The problem of compliance with formal but unproductive targets has famously been flagged by (Natsios 2010) who argues that “the U.S. aid system (and in the World Bank as well)... ignores a central principle of development theory—that those development programs that are most precisely and easily measured are the least transformational, and those programs that are most transformational are the least measurable.” (Natsios 2010, 1)

Are such distortions inevitable when using quantitative indicators for incentive purposes – or can types of indicator use minimize the risk (Table 2)?

Table 2: “Targets”, “Rankings” and “Intelligence” compared

Type of Application / Use	What it Involves	How it aims to enhance performance	How it can obstruct performance
Targets	Using numbers to set and monitor minimum thresholds of performance	Concentrate attention on improving performance in a limited number of priority areas	Can produce ratchet effects, threshold effects, output distortions
Rankings	Using numbers to compare performance of different units	Encourage “sweating and stretching” to raise overall performance, avoiding ratchet effects by focusing on relative performance among rivals	Can produce threshold effects (where ranking is categorical) and output distortions
Intelligence	Using numbers as background information for choice by users or for policy change or management intervention	Encourage informed choice or developing learning capacity and diagnostic power by adding knowledge about performance, avoiding ratchet effects, threshold effects, and output distortion from gaming behavior	Can produce ambiguity, complexity, and fragility and may be ignored by key players, especially service users

Source: Based on (Hood 2012)^{xii}

4. Must TRDM approaches be completely agnostic about “upstream” public sector systems and design?

The World Bank problem-solving approach argues that while delivery or final results is clearly the ambition of the new approach, the timescale for improved results is often longer than donor or political attention spans and so output performance measures need to be accompanied by proxies which measure system improvements further upstream (**Error! Reference source not found.**).

The “agnosticism” concerning the arrangements that will deliver the improved results is thus limited – they might be loose, but there are some priors concerning how the results are to be delivered and sustained. These upstream preferences are inevitably based on many assumptions concerning their association with downstream results (PEFA Secretariat 2009; Reid 2008; Global Integrity 2010). (Andrews et al. 2012, 6) argue that such assumptions are in effect a Trojan Horse, hiding the reappearance of best practices: “...Public Expenditure and Financial Accountability (PEFA) indicators” are essentially a disguised return to a best practice approach since they “focus developing countries on conforming with characteristics ostensibly reflecting “good international practices ... critical ... to achieve sound public financial management”. Thus the approach runs the risk of again encouraging a “Washington Consensus”-equivalent approach to public management reform, implying a simplistic transfer of OECD institutional arrangements to other settings which (Pritchett and Woolcock 2004) caricature as a “one-size fits all” approach of “skipping straight to Denmark”.

This raises the fundamental question of without some preferences concerning the upstream institutional and managerial arrangements which will achieve the improved delivery, do TRDM approaches run the risk of encouraging an uncoordinated set of sector or agency-specific

reforms?

Box 4: Implicit propositions concerning preferred upstream arrangements within the World Bank public sector management approach

The Bank’s Approach argues that, while improvements in delivery performance may take some considerable time to appear, improvements in foundational management systems are proxies of results to come – and are key to sustaining delivery improvements:

- 1) *Foundations for planning and approving the annual budget and work program:*
 - a) A good budget classification (allowing funds available to be allocated on the basis of administrative units, economic purpose and functions or programs).
 - b) A multi-year orientation, in other words a widespread recognition that deferring problems to the next year (or the next administration or management team) is unsustainable – noting that the exact form of this multi-year approach can vary significantly and there is a significant risk of ritualism in which the medium term perspective is provided on paper, but is not reflected in the mindset of senior staff.
 - c) A process for preparing the budget that is seen to be reasonable and during which the views of the spending departments are recognized.
 - d) The budget needs a degree of flexibility, for example the wage bill or other earmarks does not crowd out investment or other important expenditures.
- 2) *Foundations for implementing the annual budget:*
 - a) Confidence on the part of the spending units that they will get the funds that were budgeted.
 - b) Good recording and management of cash balances, debt and guarantees to prevent unwelcome end of year surprises.
 - c) Effective payroll controls that minimize the usual sins of ghost employees and double-dipping, with salary payments which are made on time.
 - d) Competition, value for money and controls in procurement that encourage performance and does not just focus on fiduciary controls.
- 3) *Foundations of accountability:*
 - a) Reasonably comprehensive internal and external audit.
 - b) Ensuring that the public have access to key financial/fiscal information and performance information.

II. Summary

Approaches for “targeting the results, diagnosing the means” are a welcome breath of fresh air in a difficult domain. This new focus on measurable public sector delivery improvements combined with agnosticism about how they are to be achieved is clearly in tune with the current result’s focus within the international development community. It squarely addresses many of the challenges recognized by practitioners in previous approaches. However, we still have remarkably little hard evidence on which to base a robust assessment of the effectiveness of the approach. It looks and feels encouraging – but in reality, all we know right now is that it is a significant departure from plan A. We do not know that, as a plan B, it is a success. As such, it will be very important to address the first question: ***How can we build research and evaluation into this broad family of results-based reform approaches?***

As noted above, there is a significant degree of difference within the TRDM approaches that could be explored within any overall evaluation of these approaches. Such tactical questions include:

1. ***Are “Delivery Units” or equivalent important for “anchoring” reforms?***
2. ***Does defining the performance problem at the local level, with minimal top-down involvement, result in a focus on small scale changes?***
3. ***What is the impact of a reliance on quantitative performance targets?***
4. ***Must these approaches always be completely agnostic about “upstream” public sector systems and design?***

Finally, but not insignificantly, there is the question of the political economy within donor organizations. Recent discussion within the World Bank highlights the degree to which the Bank’s new results-based lending instrument (Program for Results)^{xiii} faces implementation challenges resulting from interpretations of the Bank’s operational guidelines, which arguably push the new lending instrument back into the mold of the best practice “infrastructure template”. The final question to pursue is: ***How can donors be motivated to experiment with new modalities in an area where there is a long tradition of focusing on the best practices to be transferred rather than the delivery problems that are to be solved?***

Annex 1: “Last mile” service delivery arrangements

The key mechanisms and their drivers^{xiv}		Where is this used	What change levers does this give government?	Implied PSM capacities	Main ideological criticisms of the model	When it does not work, why does it not work
Professional Group Driven	Trust & Altruism: <ul style="list-style-type: none"> Reliance on professional standard-setting and self-regulation 	<ul style="list-style-type: none"> Health services and education have traditionally been managed on this basis 	<ul style="list-style-type: none"> Harness the public trust in professional groups Gives government strong insights into capacity gaps 	<ul style="list-style-type: none"> Complex negotiating machinery with professional groups Management of pay relativities 	<ul style="list-style-type: none"> Assumption that public servants are intrinsically motivated 	<ul style="list-style-type: none"> Professional self-interest can be redescribed as professional conclusion The model can reward failure through more resources to weak spots
	Hierarchy & “Intelligence” ^{xv} : <ul style="list-style-type: none"> Performance-informed budgeting Program budget structure Requirement to report on performance expectations in budget and describe results in entity reports 	<ul style="list-style-type: none"> Program budgeting in many settings (Russian Federation for example) 	<ul style="list-style-type: none"> Tends towards traditional use of career incentives (moving staff etc.) 	<ul style="list-style-type: none"> Strong M&E capacity Soft forms of performance management – generally limited use of performance-related pay (PRP) 	<ul style="list-style-type: none"> Assumption that public servants are intrinsically motivated 	<ul style="list-style-type: none"> Weak model – can lead to business as usual Gaming (all forms)
Senior Administrator Driven	Hierarchy & Targets: <ul style="list-style-type: none"> Performance-driven budgeting Program budget structure Requirement to report on performance expectations in budget and on results in entity reports 	<ul style="list-style-type: none"> No Child Left Behind legislation in the US UK NHS reforms 	<ul style="list-style-type: none"> Tougher use of career incentives including dismissal 	<ul style="list-style-type: none"> Strong M&E capacity Strong performance management regimes for staff – often extensive use pay flexibility and PRP Ability to manage against union opposition 	<ul style="list-style-type: none"> Assumption that public servants are extrinsically motivated by rewards 	<ul style="list-style-type: none"> Gaming (all forms) Loss of political nerve
	Choice & Competition: <ul style="list-style-type: none"> Money follows choice Supply-side flexibility Management freedom 	<ul style="list-style-type: none"> Charter schools 	<ul style="list-style-type: none"> Close institutions, allow failure 	<ul style="list-style-type: none"> Ability to move staff and resources away from failing institutions 	<ul style="list-style-type: none"> Assumption that most public servants are intrinsically motivated but that agency heads are extrinsically motivated by career rewards 	<ul style="list-style-type: none"> Gaming (primarily output distortion) Loss of political nerve Inability to define a public sector “failure regime”
Citizens/Service User Driven	Voice and Public Ranking: <ul style="list-style-type: none"> For entities or subnationals Naming & shaming in media High profile removal of ‘failing’ CEOs Name the best 	<ul style="list-style-type: none"> Citizen scorecards in the Philippines 	<ul style="list-style-type: none"> Reputation 	<ul style="list-style-type: none"> Strong M&E capacity with defensible and easily understood metrics 	<ul style="list-style-type: none"> Assumption that most public servants are intrinsically motivated but that agency heads are extrinsically motivated by reputation 	<ul style="list-style-type: none"> Gaming (primarily output distortion) Middle class capture Misunderstanding of what drives reputation in some communities

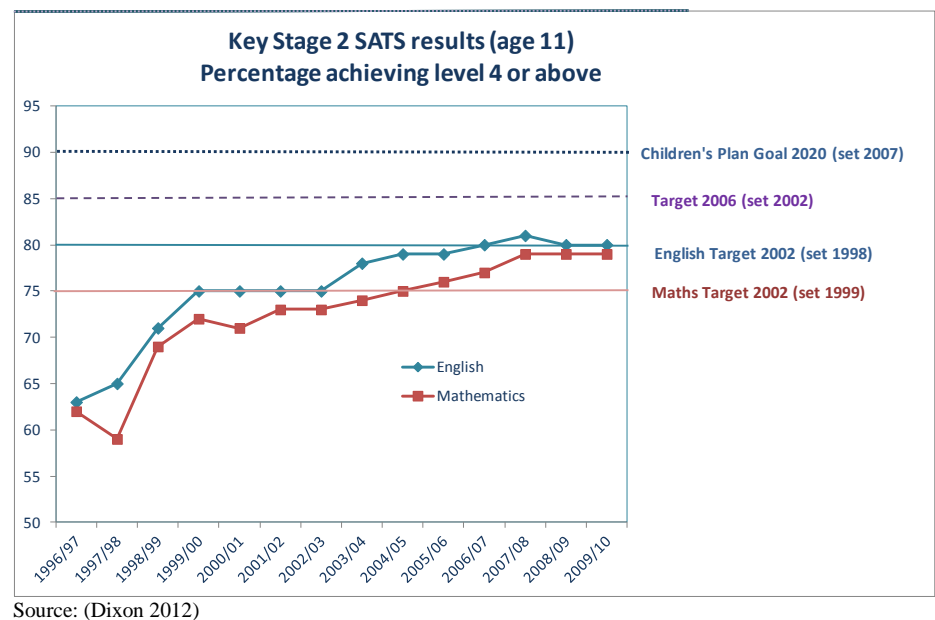
Annex 2: Delivery Units

A. The UK model

The original model for the Delivery Unit was undoubtedly that in the UK. Tony Blair’s Labour government came to power in 1997 after 18 years of Conservative rule. It was elected on strong promises of improvements to public services, especially health and education, reflecting strong public concerns about long hospital waiting lists and that less than 60 % of 11-year-olds were reaching acceptable standards of English and mathematics. Prime Minister Blair’s first term (1997-2001) was marked by a strong emphasis on performance with, inter alia, the Standards and Effectiveness Unit headed by Michael Barber established in the Department for Education, and the introduction of

Public Service Agreements (PSAs) across government, defining the performance standards that departments were to achieve. The education results were promising **(Error! Reference source not found.)** but hospital waiting lists remained long, as public opinion of the health service worsened. PM Blair made a famous ‘scars on my back’ speech in 1999 on the difficulty of public service reform.

Figure 5: Examination results in the UK (1996-2009)



In consequence, PM Blair created the PM’s Delivery Unit (PMDU) at the outset of his second term (he had comfortably won the 2001 General Election with a huge overall majority). To place this in context, the machinery of government in the UK is relatively malleable – initiatives and new organizations can be created and abolished readily, with no need for legislation. The PMDU was comprised of 40-50 staff with a mixture of civil servants, consultants, and front-line workers. It was identified strongly with the PM personally and had the remit to deliver on a handful of indicators covering Health, Education, Criminal Justice, Transport, and others. Over time, the underlying performance framework for the delivery unit changed. Cross-cutting PSAs were introduced in 2007. The approach was to set clear and ambitious targets for key service improvements, holding ministers personally accountable, while providing support through PMDU expertise and methodology. The PMDU worked with departments to agree ‘trajectories’

to meet their targets and appropriate indicators by which to judge success. ‘Stocktaking’ meetings were attended by the PM, departmental ministers and officials.

There were undoubtedly results. Hospital waiting lists in England (with high-stakes targets) dropped while those in Scotland (without such targets) did not. However, attribution is complicated as public expenditures on health care increased by over 30% (as proportion of GDP) during PM Blair’s second term. Additionally, there was some controversy concerning the metrics. The percentage of 16-year-olds gaining 5 or more GCSEs (a standard examination) at above grade C in England and Wales rose significantly, at the same time the OECD-PISA Science, Math and Reading scores of 15-year-olds fell (Dixon 2012).

B. Looking outside of the UK

In 2008, Governor O’Malley of the *US state of Maryland* created a Delivery Unit to work with state agencies to align state and federal resources around 15 strategic goals to improve the quality of life in Maryland. The goals are broadly categorized into four key areas – skills, security, sustainability, and health. The unit is comprised of around five full time staff. At bi-weekly meetings, State managers meet with the Governor and his executive staff to report and answer questions on agency performance and priority initiatives. Each week a comprehensive executive briefing is prepared for each agency that highlights areas of concern. Briefings are based on key performance indicators from a customized data template submitted biweekly by participating agencies.

In *Indonesia*, the President’s Delivery Unit for Development Monitoring and Oversight (UKP4), located in the Vice President’s Office, was established in 2009 as a temporary body (its mandate expires in 2014). It was initially focused on helping the President fulfill campaign promises (first 100 days in office) but subsequently was refocused on delivery of the 11 major priorities of the National Development Medium Term Plan. The Head has the position of a minister and is directly responsible to the President, with responsibility for monitoring achievement of the 11 national development goals, “debottlenecking” (the analysis, coordination, and facilitation to unravel the problems that occur in implementation) and the operation of a situation room within the Presidency to coordinate real-time decision making. The focus is on easily measureable, quantifiable and verifiable indicators (usually physical outputs).

The *Malaysia* Performance Management & Delivery Unit (PEMANDU) was formally established on September 16, 2009 as a unit under the Prime Minister’s Department. PEMANDU’s initial role and objective was to oversee implementation and assess progress of the Governance Transformation Program, facilitate as well as support delivery of both the National and Ministerial Key Results Areas. The government subsequently added a responsibility for its Economic Transformation Program. While responsibility for delivery of the outcomes ultimately rests with the respective ministries, PEMANDU was tasked with catalyzing bold changes in public sector delivery, supporting the ministries in the delivery planning process and providing an independent view of performance and progress to the Prime Minister and ministers. It set

performance targets in reducing crime and corruption, improving learning outcomes, reducing poverty, improving rural infrastructure and urban transport.

PEMANDU used an innovative ‘laboratory’ method which brought together civil servants and other stakeholders to identify targets and conduct the detailed planning through participation that would be needed to achieve results. The government committed substantial resources to the PEMANDU initiatives: US\$3.07 billion was budgeted in 2011. In education alone, supporting the targets on literacy, numeracy and pre-school enrolment included:

- RM10 million for pre-school building (\$1 = RM3.11);
- Pre-school fees for 14,000 children;
- 500 grants for pre-schools @ RM10K per grant;
- RM 700K for each of 52 ‘high-performing’ schools; and
- RM7500 for the Heads of those schools, and RM900 for each of the teachers.

After two years of operation, the government reported dramatic improvements, including a reduction in overall crime of 11.1% against the government’s initial target of 5%. Targets included an improvement in Transparency International’s (TI) Corruption Perceptions Index from 4.4 to 4.9. The index released by TI in December 2012 exactly matched the target which the government had set two years earlier.

The government appointed an international review panel, whose members included an Australian Public Service Commissioner and then IMF Resident Representative. It was generally positive in its assessment (Mccourt 2012). Somewhat echoing debates in the UK, there was some controversy about the validity of some of the metrics, and the close identification of the Delivery Unit with the Prime Minister was seen as both a source of political authority and a potential challenge to sustainability. The review panel was concerned that GTP is “vested in the Prime Minister alone.”

From 2009 through 2012, the Office of Tony Blair was involved in an effort to set up a service delivery unit in the Prime Minister’s Office in *Kuwait* exactly along the lines of the one in his Prime Minister’s Office. In spite of the investment of considerable resources, the initiative was unsuccessful as the unit lacked close engagement with Prime Minister who had somewhat mixed views on the utility of the unit and of performance improvements more generally.

The equivalent unit in *Chile* was established in 2010 by President Piñera with the objective of improving performance with a new focus on delivery. The Presidential Delivery Unit (Unidad Presidencial de Gestión del Cumplimiento) at the Ministry of the Presidency, aims to fast-track delivery improvements in seven priority areas: growth; employment; public security; education; health; poverty; democracy, decentralization and state modernization. It also has responsibility for accelerating the post-earthquake reconstruction program. The Unit coordinates with the Budget Directorate of the Ministry of Finance, the Secretary General, the Undersecretaries of the Interior and Regional Development, and the Presidential Advisory team. While the unit has made an important contribution to the identification of policy priorities, the development of robust

indicators to monitor them and in supporting ministries in improving the quality of policy, it is still early to assess its impact on improving public service delivery.

These units all have the hallmarks of the UK PMDU and were established with assistance from McKinsey and Co. – the consulting firm where Michael Barber initially worked following his departure from government in the UK.

There have also been attempted replications in *Africa*, under the auspices of the Africa Governance Initiative. Information about them is limited. There are mixed reports about these attempts (Scharff 2012). Based on the PEMANDU experience, the Government of Tanzania is establishing a transformational Big Results Now! program to identify and resolve constraints to results delivery in six priority areas. The Government of Tanzania plans to commission the expertise of the Government of Malaysia to design and implement the program, run the labs and establish a President’s Delivery Unit. A somewhat similar initiative in Rwanda used existing strong central units rather than build a new Delivery Unit (Annabel Jackson Associates 2009).

There has also been some replication of the approach at the sector level. The establishment of the “Programme Monitoring and Implementation unit” for the Punjab education sector has followed a path recognizably similar to that of the cross-cutting, multi-sector delivery unit models (Barber 2013).

In sum, while there has been significant replication of the UK Delivery Unit (at least rhetorically) and while there are some performance successes claimed for them, as with almost all public management innovations, attributing causality for any associated performance gains is challenging as there are always other performance-enhancing reforms underway at the same time: restructuring civil service remuneration and careers, performance-based contracts with service-providing entities, and increasing delegation and autonomy to more departments to enable them to run on more business-like lines. Monitoring of performance is also generally increasing in all settings, with or without delivery units.

C. Commonalities

Delivery units have a distinctive role – generally undertaking all or a combination of the following five functions:

1. Focusing political pressure for results through progress-chasing on behalf of the head of government;
2. Providing a simple and direct monitoring mechanism for key government priorities;
3. Signaling key government delivery priorities within and outside of the public sector;
4. Providing a clear signal, at least internally, that government is holding ministers and senior staff to account for delivering the government’s key strategic priorities; and
5. Supporting innovation, coordination by various ministries, and providing a forum for problem solving when needed.

By their nature, creating a delivery unit reflects political impatience – a concern that the current governmental performance system is not delivering results quickly enough or not in the desired areas.

Delivery units are generally comprised of a small cadre of highly skilled staff, often with a combination drawn from the public and private sectors, that seeks to work in partnership with ministries/agencies. They all have direct access to the political leadership in order to initiate authoritative and binding problem-solving meetings of senior policy makers and senior civil servants. They focus on a limited number of explicit, public government priorities and establish a light, nimble data collection and reporting system at the apex of a system of regular performance monitoring to ensure that responsible ministers maintain a continual focus on the objectives.

References

- Acemoglu, D., Johnson, S. & Robinson, J. A. 2001. “The Colonial Origins of Comparative Development: An Empirical Investigation.” *American Economic Review* 91: 1369-1401.
- Acemoglu, D. & Robinson, J. A. 2012. *Why Nations Fail: The Origins of Power, Prosperity and Poverty*. New York: Crown Publishers.
- Andrews, M. 2009. *Isomorphism and the Limits to African Public Financial Management Reform*. Cambridge, MA: John F. Kennedy School of Government - Harvard University.
- Andrews, M. 2013a. “Building State Capability in Mozambique’s Judicial Sector.” In *Governance Reform in International Development*, ed. Andrews, M. Cambridge, MA: Matt Andrews.
- Andrews, M. 2013b. *The Limitations of Institutional Reform in Development*. New York: Cambridge University Press.
- Andrews, M. 2013c. “What Is Governance?” In *Governance reform in international development*, ed Andrews, M. Cambridge MA: Matt Andrews.
- Andrews, M., Pritchett, L. & Woolcock, M. 2012. “Escaping Capability Traps through Problem-Driven Iterative Adaptation (PDIA).” Working Paper 299, Center for Global Development, Washington, DC.
- Annabel Jackson Associates. 2009. *Africa Governance Initiative: Rwanda Pilot Evaluation*. London: Gatsby Charitable Foundation.
- Banerjee, A. & Pande, R. 2009. *Parochial Politics: Ethnic Preferences and Politician Corruption*. Cambridge MA: Harvard University.
- Barber, M. 2008. *An Instruction to Deliver: Tony Blair, the Public Services and the Challenge of Delivery*. London: Politics.
- Barber, M. 2013. *The Good News from Pakistan*. London: Reform.
- Barber, M., Kihn, P. & Moffit, A. 2011a. *Deliverology: From Idea to Implementation*. Washington, DC: McKinney and Co.
- Barber, M., Moffit, A. & Kihn, P. 2011b. *Deliverology 101: A Field Guide for Educational Leaders*. Thousand Oaks: Sage.
- Bevan, G. 2012. *The Challenge of Designing ‘Good Enough’ Performance Measures & Results Framework*. London: London School of Economics.
- Bevan, G. & Hamblin, R. 2009. “Hitting and Missing Targets by Ambulance Services for Emergency Calls: Effects of Different Systems of Performance Measurement within the UK.” *Journal of the Royal Statistical Society* 172 (1): 161-190.
- Bevan, G. & Hood, C. 2005. *What’s Measured Is What Matters: Targets and Gaming in the English Public Health Care System*. London: Economic and Social Research Council.
- Birdsall, N. & Savedoff, W. 2010. *Cash on Delivery: A New Approach to Foreign Aid*. Washington, DC: Center for Global Development.
- Blum, J., Manning, N. & Srivastava, V. 2012. *Public Sector Management Reform: Toward a Problem-Solving Approach*. Washington, DC: World Bank.
- Booth, D. 2012. *Development as a Collective Action Problem: Addressing the Real Challenges of African Governance*. London: Overseas Development Institute.
- Brown, E. & Smith, M. 2013. “Cartac Discusses PFM Reform Strategies and State Enterprises.” In *Public Financial Management Blog*, ed. IMF. Washington, DC: IMF.
- Di Maggio, P. J. & Powell, W. W. 1983. “The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields.” *American Sociological Review* 48(April): 147-160.
- Dixon, R. 2012. “The Prime Minister’s Delivery Unit 2001 – 2010.” Presentation to the World Bank, Washington, DC.
- Dollar, D. & Kraay, A. 2003. *Institutions, Trade, and Growth: Revisiting the Evidence*. Washington, DC, World Bank.
- Easterley, W. 2008. *The White Man’s Burden: Why the West’s Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. Oxford: Oxford University Press.
- Evans, P. 2004. “Development as Institutional Change: The Pitfalls of Monocropping and the Potentials of Deliberation.” *Studies in Comparative Institutional Development* 38 (4): 30-52.

- Frumkin, P. & Galaskiewicz, J. 2004. “Institutional Isomorphism and Public Sector Organizations.” *Journal of Public Administration Research and Theory* 14 (3): 283-307.
- Glewwe, P., Ilias, N. & Kremer, M. 2010. “Teacher Incentives.” *American Economic Journal: Applied Economics* 2 (3): 205-227.
- Global Integrity. 2010. *Global Integrity Index: 2009*. Washington, DC: Global Integrity
<http://report.globalintegrity.org/globalIndex.cfm>.
- Gupta, S. & Verhoeven, M. 2001. “The Efficiency of Government Expenditure: Experiences from Africa.” *Journal of Policy Modeling* 23: 433-67.
- Hausmann, R., Pritchett, L. & Rodrik, D. 2005. “Growth Accelerations.” *Journal of Economic Growth* 10 (4): 303-329.
- Heifetz, R. A. 1994. *Leadership without Easy Answers*. Cambridge, MA: Harvard University Press.
- Henderson, J., Hulme, D., Jalilian, H. & Phillips, R. 2003. *Bureaucratic Effects: ‘Weberian’ State Structures and Poverty Reduction*. Manchester, UK: Chronic Poverty Research Centre.
- Hood, C. 2006. “Gaming in Targetworld: The Targets Approach to Managing British Public Services.” *Public Administration Review* 66 (4): 515-521.
- Hood, C. 2010. *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*. Princeton, NJ: Princeton University Press.
- Hood, C. 2012. “Public Management by Numbers as a Performance-Enhancing Drug: Two Hypotheses.” *Public Administration Review* 72: 4.
- Hughes, R. 2010. “Case Study 5.1: Using Performance Information in the United Kingdom.” In *Results, Performance Budgeting and Trust in Government*, ed. Arizti, P., Brumby, J., Manning, N., Senderowitsch, R. & Thomas, T, 167-179. Washington, DC: World Bank.
- Independent Evaluation Group. 2008. *Public Sector Reform: What Works and Why?* Washington, DC: World Bank.
- Joshi, A. & Houtzager, P. P. 2012. “Widgets or Watchdogs?” *Public Management Review* 14 (2): 145-162.
- Knack, S. & Keefer, P. 1995. “Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures.” *Economics and Politics* 7 (3): 207-27.
- Le Grand, J. 2007. *The Other Invisible Hand: Delivering Public Services through Choice and Competition*. Princeton: Princeton University Press.
- Mansuri, G. & Rao, V., Eds. 2012. *Localizing Development: Does Participation Work?* Washington, DC: World Bank.
- Mauro, P. 1995. “Corruption and Growth.” *Quarterly Journal of Economics* 110 (3): 681-712.
- Mccourt, W. 2012. “Reconciling Top-Down and Bottom-Up: Electoral Competition and Service Delivery in Malaysia.” *World Development* 40 (11): 2329-41.
- Moynihan, D. P. 2006. “Managing for Results in State Government: Evaluating a Decade of Reform.” *Public Administration Review* 66: 77-89.
- Natsios, A. 2010. *The Clash of the Counter-Bureaucracy and Development*. Washington, DC: Center for Global Development.
- North, D. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
- North, D. C., Wallis, J. J., Webb, S. B. & Weingast, B. R. 2007. “Limited Access Orders in the Developing World: A New Approach to the Problems of Development.” Policy Research Working Paper 4359, World Bank, Washington, DC.
- PEFA Secretariat. 2009. *Public Expenditure and Financial Accountability: Public Financial Management Performance Measurement Framework*. Washington, DC: PEFA Secretariat, World Bank
<http://www.pefa.org/>.
- Pollitt, C. & Dan, S. 2011. “The Impacts of the New Public Management in Europe: A Meta-Analysis.” COCOPS Working Paper 3, European Commission, Brussels.
- Pritchett, L. & Woolcock, M. 2004. “Solutions When the Solution Is the Problem: Arraying the Disarray in Development.” *World Development* 32 (2): 191-212.
- PRMPS. 2012a. “Indicators of the Strength of Public Management Systems: A Key Part of the Public Sector Management Results Story.” PRMPS Discussion Paper Draft April 22, 2012, World Bank, Washington, DC.

- PRMPS. 2012b. “Indicators of the Strength of Public Management Systems: Technical Annex.” Draft April 22, 2012, Washington, DC, World Bank.
- Radin, B. A. 2006. *Challenging the Performance Movement: Accountability Complexity and Democratic Values*. Washington, DC: Georgetown University Press.
- Reid, G. 2008. *Actionable Governance Indicators – Concepts and Measurements*. Washington, DC: World Bank.
- Rodrik, D. 2008. *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth*. Princeton, NJ: Princeton University Press.
- Rodrik, D., Subramanian, A. & Trebbi, F. 2004. “Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development.” *Journal of Economic Growth* 9: 2.
- Sacks, A. 2011. “The Antecedents of Approval of the Incumbent Government and Trust in Government in Sub-Saharan Africa, Latin America and Six Arab Countries.” *Public Sector Management Companion Notes*. Washington, DC: World Bank.
- Scharff, M. 2012. *Delivering on a Presidential Agenda: Sierra Leone’s Strategy and Policy Unit, 2010 – 2011*. Princeton: Princeton University.
- Srivastava, V. & Larizza, M. 2012. *Working with the Grain for Reforming the Public Service: A Live Example from Sierra Leone*. Washington, DC: World Bank.
- World Bank. 2000. *Reforming Public Institutions and Strengthening Governance: A World Bank Strategy*. Washington, DC: World Bank.
- World Bank. 2003. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank.
- World Bank. 2012a. *Program-for-Results Information Document (Punjab Governance Reforms for Service Delivery)*. Washington, DC: World Bank.
- World Bank. 2012b. *Program-for-Results: An Overview*. Washington, DC: World Bank.
- World Bank. 2012c. *The World Bank’s Approach to Public Sector Management 2011-2020: Better Results from Public Sector Institutions*. Washington, DC: World Bank.

-
- ⁱ Source: World Development Indicators. In 2011, the under-5-year old mortality rate in Nigeria was 124.1 deaths per 1000 births, compared to 77 deaths per 1000 births in Ethiopia. In 2010, health expenditures per capita were 62.8 USD in Nigeria, compared to 15.7 USD in Ethiopia.
- ⁱⁱ Source: World Development Indicators. In 2010, the total youth literacy rate (% of people ages 15-24) in Benin was 50.5 percent, whereas it was 80.8 percent in Ghana.
- ⁱⁱⁱ Source: World Development Indicators. In 2011, the infant mortality rate in Sweden was 2.2 children per 1000 births, compared to 6.4 children in the US. In 2010, per capita health expenditure in Sweden was 4710.4 USD, compared to 8361.7 USD in the US.
- ^{iv} In economics, Dani Rodrik talks of ‘one economics, many recipes’ (Rodrik 2008), Bill Easterly talks of development practitioners moving ‘from planners to searchers’ (Easterley 2008). In institutional reform, (Heifetz 1994) has long emphasized ‘adaptive versus technical problems’ and (Evans 2004) has warned about institutional “monocropping” as an alternative to “deliberation”. (Blum et al. 2012; World Bank 2012c; World Bank 2000) review how these are translating into a more agnostic approach in public sector management reforms. As (Booth 2012) notes: “Reformers should recognise that essential institutional *functions* for economic and social development can be fulfilled with quite varied institutional *forms*” (Booth 2012, 6).
- ^v The assumption that public sector organizations and institutional arrangements should look similar even if they achieve very different results in different settings is often referred to as “isomorphism” (Di Maggio and Powell 1983). (Andrews 2009) argues that development agencies including the World Bank have an isomorphic mindset, generally proposing reforms that standardize institutional arrangements between settings. Isomorphism stems from three pressures – coercive (funders and managers, and wider society, expects a particular organization to look like organizations doing similar things in other settings); mimetic (when managers and reformers are not sure what to do, then restructuring to look like other organizations is understandable); and normative (when the dominant professional culture, accountancy for example, shares similar assumptions about the right way to do something) (Di Maggio and Powell 1983). Of these, perhaps the strongest pressure for

isomorphism is a response to uncertainty when there is no obvious first best way forward (Frumkin and Galaskiewicz 2004). Isomorphism can equally apply to reform tools as (Joshi and Houtzager 2012) note in identifying the temptation to use reforms ‘widgets’ – a method which has emerged in one country to meet one set of needs, but which may not succeed in a different context.

- vi The general case is summed up well by (Henderson et al. 2003) when they note that: “there is in general a strong relation between the competence and effectiveness of public bureaucracies and their consequences for poverty reduction. While it is important to recognise that correlations are not the same as causal connections and that in the social world the latter rarely, if ever, can be empirically ‘proved’, we suggest that given a solid and sustained record of economic growth, the balance of presumption must be that the bureaucratic quality of public institutions in a given country is decisive for that country’s ability to reduce poverty” (Henderson et al. 2003, p.15). However, the specifics of public sector management reform are significantly under-researched – leaving much hard evidence pointing to rather general conclusions that the institutions that have been shown to matter for economic development are largely restricted to those that protect the returns to private investment, in particular property rights and the rule of law.

On this general level, it has become commonplace to recognize that “institutions matter” (North 1990) for economic development, with cross-country empirics relating better institutional quality to higher levels of per capita income and greater economic growth (Mauro 1995; Knack and Keefer 1995; Acemoglu et al. 2001; Dollar and Kraay 2003; Rodrik et al. 2004). A foundational level of institutional quality in relation to property rights and the rule of law appears to be necessary for sustained economic growth (Acemoglu et al. 2001) and (Rodrik et al. 2004) – but beyond that, it could be that institutions are an outcome of economic development as richer societies demand better governance structures. This ambiguity is underscored by findings that demonstrate that growth accelerations are often not preceded by or tied to major changes in core public sector institutional arrangements (Hausmann et al. 2005). The growth experiences of China after the late 1970s and South Korea from the early 1960s provide two such examples.

While the past decade has seen a tremendous growth of experimental studies on the effectiveness of management reforms in sectors – such as teacher or health worker incentives – for learning or health outcomes – no comparable revolution has happened in the knowledge on how to reform upstream public sector institutions. Consequently, there is relatively limited scientific evidence about what matters most for improving public sector performance. There are many possible reasons why research on PSM reform in developing countries is lagging behind. They include that there are more economists than public administration scholars focusing on developing countries, and that PSM reforms are long term, complex and tough to measure, lending themselves less to rigorous evaluation. Unlike deworming pills, a Medium Term Expenditure Framework cannot be randomized. This is not to say that the field of PSM research is without advances – but compared to other policy domains there is relatively little evidence about what matters most in improving public sector performance, in particular in developing countries. While the past decade has seen a tremendous growth of experimental studies on the effectiveness of management reforms in sectors – such as teacher or health worker incentives – for learning or health outcomes (See for example (Glewwe et al. 2010) for a comprehensive review of the rapidly growing rigorous impact evaluation literature on teacher incentives, or (Mansuri and Rao 2012) for a comprehensive review of the research on different forms of citizen participation.) – no comparable revolution has happened in the knowledge on how to reform upstream public sector institutions.

- vii (Moynihan 2006) documents how this imbalance has led to a political logic of *partially* implementing New Public Management reforms in the United States at the state level. Politicians often did strengthen public managers’ accountability for results – which is a popular stance to take. But they failed in giving managers the necessary freedom for managing staff and money to achieve these results – as this would have implied confronting powerful vested interests, such as public sector unions.

- viii It is important to note that the 60’s notion of “capacity” was a narrow concept focusing on the lack of technical ability to perform a task, in contrast to more recent, broader ideas of capacity which comprise political commitment and institutional design.

- ix Such bodies include the Ministry of Finance and the offices that support the head of government.

- x Over a five year period, the UK moved from reporting on 600 to 30 main ‘priorities’, although the burden of reporting by front line services remained high. Learning from that experience, Malaysia’s Delivery Unit started

out with only six. In Indonesia, Thailand and Malaysia the initial systems often had many more, even thousands, of mainly output indicators.

- ^{xi} (Hughes 2010) details the number of performance targets required for front line health and education specialists in the UK under the Public Service Agreement framework. There was widespread concern expressed in academic and media articles about how this was increasing the transaction costs of public service delivery without necessarily achieving the desired performance improvements.
- ^{xii} Hood suggested that the magnitude of these gaming risks is significantly shaped by the culture of those being measured: “hierarchist”, “egalitarian”, “individualist”, or “fatalist”
- ^{xiii} See World Bank. 2011. *Program-for-Results: An Overview*.
http://siteresources.worldbank.org/EXTRESLENDING/Resources/7514725-1313522321940/PforR_Overview_12.2011.pdf
- ^{xiv} Developed from (Bevan 2012; Le Grand 2007)
- ^{xv} “Intelligence” refers to the general provision of performance information but with no particular incentives attached to it