

Managing a Sustainable Results Based Management (RBM) System

This note presents a framework for thinking about public sector results based management (RBM) systems, with a particular focus on the issues line agencies face in complying with mandates and directives from central agencies on monitoring and evaluating performance. It also provides five lessons learned from RBM systems of relevance for countries pursuing results based management reforms. Taking a system's view of results based management reveals a number of different approaches and techniques used across the public sector to improve results.

Background

A public sector results based management system consists of a variety of approaches, incentives, tools, and techniques that public sector entities adopt to help monitor, evaluate, and improve performance results. Many of the tools and techniques were first adopted in the private sector, and over time, have been adapted for public sector purposes, such as balanced scorecards and the total quality management framework. Results based management systems deliver value through a number of channels such as, aligning different organizations to move toward common objectives or creating an environment in which agencies learn and improve program design. The main drivers for introducing a Results Based Management system include the Ministry of Finance/Budget Office need for a basis for the allocation of funds and desire to know what agencies are buying, a desire by elected officials to hold the bureaucracy to account, the desire to encourage agencies to think about their production function and efficiency, to link their outputs to impact, and to link service quality to citizen needs (responsiveness). Generally, one can think of RBM systems designed to emphasize relatively more of one of these objectives:

- To learn;
- To steer and control; and
- To provide accountability.¹

In pursuit of one of these objectives or some combination thereof, countries often employ a number of results based management tools. These include, strategic planning, internal management information systems, benchmarking, risk analysis, or business process reengineering geared toward facilitating an internal management dialogue. Other countries may use monitoring and management scorecards,² performance agreements, citizen charters, annual reporting, and performance contracts. The objective/purpose of the system has implications for

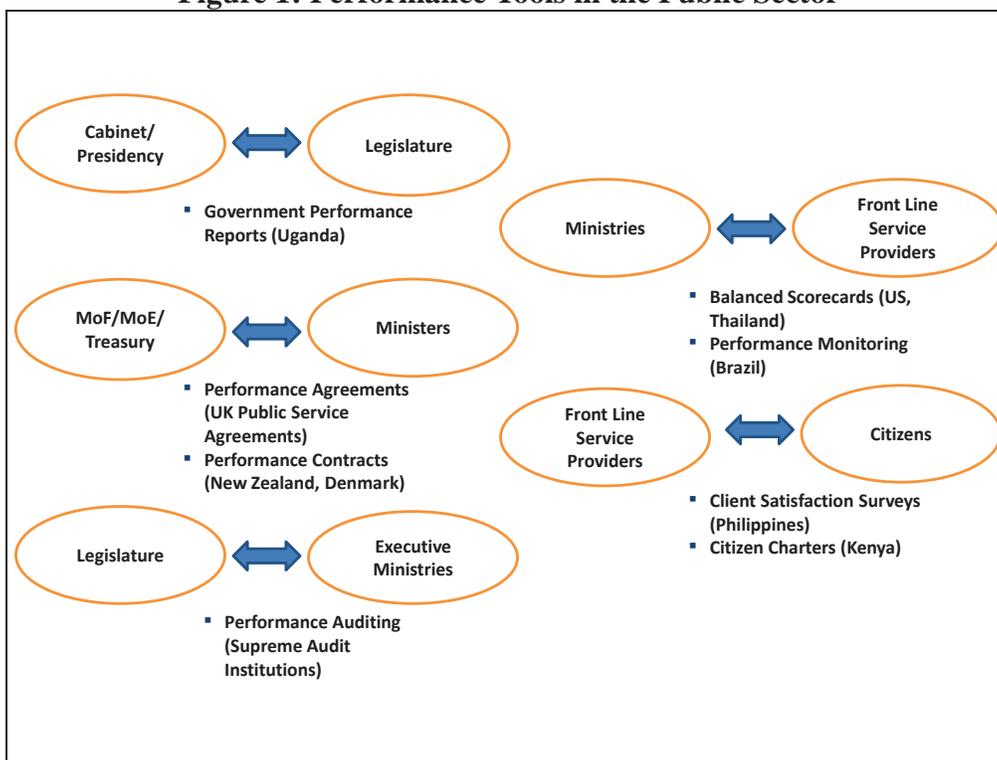
GET Notes – Recently Asked Questions Series intends to capture the knowledge and advice from individual engagements of the World Bank's Global Expert Team on Public Sector Performance (PSP GET). The views expressed in the notes are those of the authors – William Dorotinsky and Joanna Watkins - and do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

¹ Wouter Van Dooren et al, *Performance Management in the Public Sector*, (2010).

² The Balanced scorecard is a strategic performance management tool that was first developed by Art Schneiderman in 1987 and the method was further developed for practical use by Kaplan and Norton in 1996. It enables organizations to translate their vision and strategy into implementation through four perspectives: financial, customer, business process, and learning/growth.

which tools are selected and how each of them is designed –specifically in terms of the accountability arrangements surrounding the implementation of the tool. Different performance tools have been developed to meet different objectives and problems, and address different linkages and accountability arrangements among various actors within the public sector. Many of these are mutually reinforcing, creating a network of supporting arrangements. Figure 1 presents a number of tools for reinforcing the linkages between different public sector actors. As the implementation of various aspects of results based management systems are put into place, it may be helpful to pay attention to the weakest link, where the reform would most improve performance.

Figure 1: Performance Tools in the Public Sector



In public sector management, there is always a tension between central (Headquarters) control and managerial responsibility. Technology and performance monitoring often is introduced to increase central control without due attention to efficiency impacts or institutional learning. There are a plethora of examples of different RBM systems from across the globe. Some emphasize more of one objective than another and often at different levels of government. The institutional and functional variability of the systems makes it difficult to make cross-national comparisons. Notable examples include Colombia’s National System for Evaluation and Management for Results (SINERGIA), Mexico’s National Council for the Evaluation of Social Development Policies, the United States Government Performance and Results Act (GPRA) and the Management Control Division in the budget department of the Ministry of Finance in Chile.

Often accompanying these tools are a set of incentives at the organizational and individual level, which include monetary and non-monetary rewards for organization/individual performance, and delegation of managerial authority for human resources/budgeting for improved efficiency.

Generally, the measurement of results/performance of a project/program is in part determined by where you sit. As a purchaser of a service, your perspective is typically different from those of the supplier of the service. For example, as the President/Prime Minister your preferences/interests may differ substantially from those of the Ministry of Finance or Line Agency. As a regulator, your interests (e.g. compliance with the law) are typically different from those of the regulated – who are often concerned with the lightness of the regulatory burden. In some cases, parties may have interests as both purchasers and suppliers (e.g. co-operatives). Government performance is complex in part because of the variety of roles and perspectives of different actors.

RBM systems evolve over time to respond to various political demands, technological advances, and an improved understanding of how to measure and assess performance in sectors. With any RBM system, there are a number of critical limitations to be aware of – these relate to the measurability of certain activities, outputs/outcomes, over- or under representation of measured values, the use of too many indicators, lack of clarity on definitions and understanding of the performance information being produced leading to misinterpretations, and the manipulation of data (in terms of unjustifiable aggregation or disaggregation). Such risks cannot be eliminated, but only mitigated by using caution in interpreting information derived from the RBM system and responding to the inherent complexity in the public sector production chain in the design of the results based management system. Table 1 presents a few of the common risks/problems associated with performance measurement and potential solutions to mitigate such risks.

Table 1: Common Risks/Problems with RBM systems

Common Risks/Problems	Mitigation/Solutions
Over or under representation of measured values	Caution in interpretation of statistics, open dialogue, careful use of incentives
Use of too many indicators	Link indicators to a limited set of top priorities, limit reporting on indicators to appropriate levels
Lack of clarity on definitions and guidance	Issue guidance on core definitions, standardize and streamline reporting
Manipulation of data	Careful citation of sources, open discussion on data, use of IT systems, and use of qualitative assessment to corroborate findings
Unjustified aggregation/disaggregation	Open discussion on data, cautious interpretation of statistics

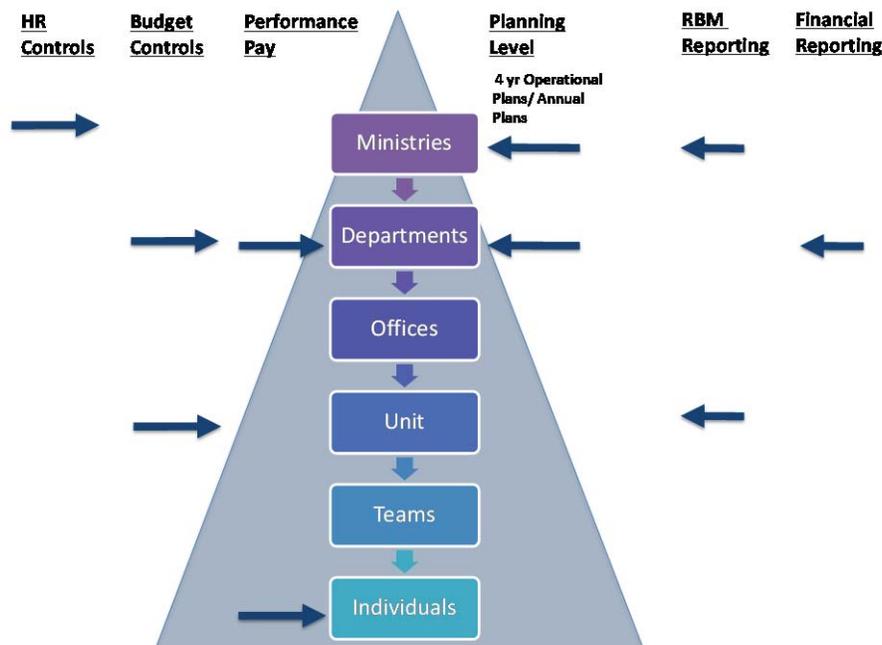
Many OECD country systems are at the stage of making publically available performance data, learning and improving performance assessment, rather than in linking the results with resource allocation. The old adage ‘you get what you measure’ holds true, and unless the measurement and reporting are well developed for a sector and stable, adding monetary incentives to less developed systems may rapidly lead to adverse outcomes and negative externalities.

Techniques for mitigating some of these risks associated with performance measurement include, publishing results at various levels of aggregation for the public (getting the right data to the right people at the right time), citizen monitoring of government projects, decoupling performance results from budget allocations, and introducing qualitative measures to supplement quantitative data, among other techniques.

A Bird’s Eye View

Commonly, results based management approaches and processes evolve over decades independent of one another. Therefore, when analyzing an RBM system from a high level perspective, it is useful to look at a handful of important control functions (e.g. HR and budgeting), incentives (such as performance pay), and reporting lines to see where in the internal administration they intend to exert influence. Ideally the design of these central systems should be aligned to reinforce one another, rather than to create fragmentation with multiple levels of accountability. The basic premise is that entities can neither improve their performance nor be held accountable if they have no authority to actually manage their own resources, as well as tools to support this authority. Figure 2 takes a few of these control functions, incentives, and reporting requirements and maps where in a hypothetical internal administration they might exert influence. For example, an entity responsible for delivering its outputs may not receive regular financial reports from a central treasury, preventing the entity from being able to actively monitor their budget (unless they develop their own parallel financial reporting system).

Figure 2: Misalignment of various systems (hypothetical)



As reflected in the diagram, the systems for HR control, budget controls, performance pay, planning, and performance reporting are not well aligned, reflecting a lack of system coherence either towards control or supporting management. Typically in British Westminster systems, the emphasis is placed on ministerial accountability, with controls, accountability, and reporting aligned at this level. Incentives are then tied to the accountability arrangements. In the figure, HR controls are at the ministerial level, while budget controls are at the ministry and unit levels. Performance pay is at the department and individual levels. Planning is done at the ministry and department level, as is reporting on RBM for the most part. Generally, countries with more mature RBM systems attempt to align some of the systems for accountability, controls, and incentives in order to establish a meaningful 'unit of accountability' in the RBM system and avoid conflicts, which can arise from disjointed controls and accountability arrangements. For the incentives in the system to work properly, both authority and accountability need to be aligned.

It is very difficult to hold managers accountable for performance if the following exists:

- Unclear organizational structure, mandate
- They were not involved in budget setting, planning, target setting
- Performance metrics, targets change annually
- Do not actually receive approved budget levels; no predictability in funds availability
- Have no influence over procurement processes for their work

- Have no influence or human resource management for their office (positions or employment)
- Do not receive continuous or regular spending reports for their units
- No regular management or performance and output reports
- Have no flexibility or discretion in resource allocation
- Have treasury/accounting systems on a cash basis, necessitating their own parallel commitment accounting
- Managers selected for technical skills, not management ability
- No managerial training, support system on performance results

Therefore, the actual impact of any performance system can be considered a function of how much it changes behavior appropriately and achieves results, its ease of use, cost to operate (time & money), and fit for purpose. In other words, the impact of an RBM system is a function of desired behavior change divided by the complexity of the system. As the complexity of the system increases, its impact decreases. If the RBM system seeks to measure everything with no selectivity, the systems may actually end up measuring and achieving nothing.

$$\text{Impact}_{(\text{real})} = \frac{\text{Positive Behavioral Change}}{\text{Complexity of System}}$$

Lessons from International Experience

Over the past decade, a wealth of information has been collected on the experiences of countries pursuing results based management reforms. The OECD’s “In Search of Results: Performance Management Practices” (1997) analyzed a wide variety of country experiences along the dimensions of objectives and approaches, institutional arrangements, performance measurement, financial management, and the reporting of performance information.³ Five important lessons are presented here based on the findings of OECD, World Bank, and other reports of relevance for countries pursuing results based management.

Lesson 1: Over-collection of performance information and setting too many targets limits the effectiveness of an RBM system.

The lessons from the UK’s experience with high level targets illustrate this point well. In connection with the 1998 Comprehensive Spending Review exercise, Public Service Agreements (PSAs) were introduced setting around 600 performance targets for roughly 35 areas of

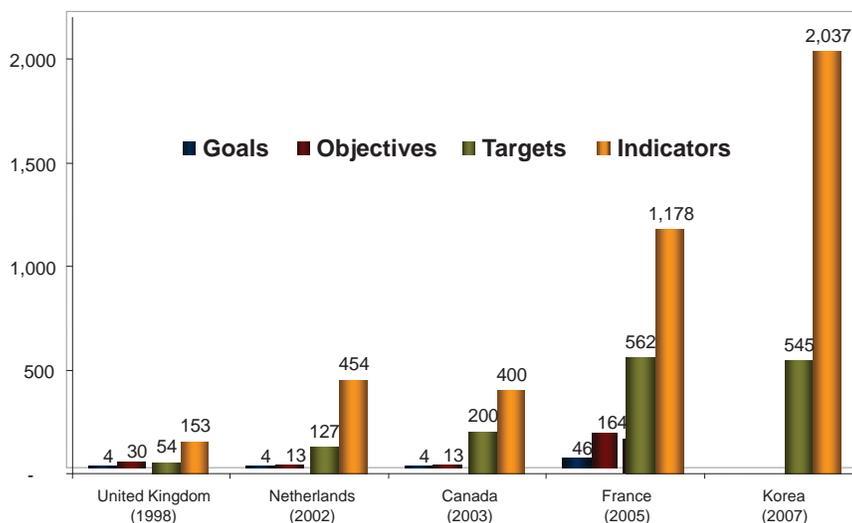
³ OECD, *In Search of Results*, PUMA, 1997.

Government. Over time these were refined and reduced, as a result of feedback from agencies. Setting the right PSAs was a challenge – in the first round only 15% of PSAs were outcome-based, and they were not matched to the budgeting cycle. Moreover, a lesson learned was that setting the PSA targets was not sufficient to achieve results without a more detailed action plan for the Ministry explaining how the PSA target would be achieved.

By the Spending Review in 2000, the number of top level PSAs were reduced to around 160, supported by a range of Service Delivery Agreements (SDAs), which included a lower level set of targets for the purpose of departmental delivery planning. In 2001, the United Kingdom established the Prime Minister’s Delivery Unit (PMDU) to help monitor progress on and strengthen the UK Government’s delivery of public service priorities through a sustained focus on the performance of key services and public sector management – including supporting the foundations of PSAs. By 2009, PSAs had been reduced to 30, signaling the major outcome-oriented priorities of government.

The evolution present in the UK experience reflects the tendency when countries first start out to want to monitor the performance of everything in the system as illustrated in Figure 3. Generally systems will go through a period of intense data collection, but over time the number of indicators/targets reported to the national level tends decreases.

Figure 3: Number of national level indicators used by central government



Source: OECD, 2007.

In the case of US State of Oregon, the “Oregon Shines II” program and evolution of the Oregon Benchmarks also reflects a similar evolution. At the start of the program, there were over 270 Oregon Benchmarks. The sheer number of indicators caused confusion and misunderstanding, and were basically ineffective for decision making purposes. After years of careful observation and refinement, Oregon narrowed its benchmarks to 90, to fit within seven major areas, which

have been universally accepted as both “meaningful and otherwise helpful.”⁴ Many countries’ RBM systems initially collect more performance information than they know what to do with or can effectively use, and over time rationalize reporting.

Lesson 2: Simply collecting performance information doesn’t necessarily lead to improvements.

Performance information is often collected in a wide variety of formats and for a number of intended users. No one system is likely to meet the information needs of all users. The usefulness of this information for any given user depends on what information is being collected, the quantity and quality of the information, how it is presented, the intended purpose, and the timeliness of the information for relevant decision-making processes (e.g. the budget or planning cycle). The non-use of performance information can stem from a misalignment in any of these areas, as well as its interaction with other inputs in decision-making (e.g. political influence). In addition, a well documented phenomenon is that of ‘gaming’ or manipulation of indicators or results for reporting purposes. Such contamination in the data undermines its credibility and efficacy for decision-making purposes. Getting users to use the performance information for decision-making is therefore contingent on a number of factors and requires careful attention to quality assurance and usability. Box 1 describes the quality assurance framework in place to support the UK Public Service Agreements.

Box 1: Improving the quality of performance information— UK (2001-2010)

From 2001-2010, at the heart of the UK’s performance management system were Public Service Agreements (PSAs), agreed between the Finance Ministry and line ministries. Introduced in 1998, their aim was to focus resources on improving outcomes for the public and to strengthen accountability for cost effective service delivery. Published alongside departments’ three-year budget allocations, PSAs specify: i) the department’s aim; ii) five to 10 supporting objectives; iii) performance targets, including a value-for-money target; and iv) standards to be maintained, monitored and reported. PSA targets were refined gradually to focus on outcomes rather than the inputs or processes. *The number of targets has been reduced, from around 400 in 1998 to 30 in the 2008-11 Spending Review.* Biannual reports are published, which provide information on spending and performance against PSA targets. While the public and parliamentary oversight of performance has generally been weak, there has been a great deal of emphasis on quality assurance of the performance measures. The Office for National Statistics provides advice to ministries and agencies on methods, and on quality assurance for statistical systems through central government, certifying appropriate indicators. The National Audit Office, who has created a Directorate of Performance Measurement to co-ordinate work on performance measurement in financial audits, provides advice, training and reviews of central government bodies’ governance (and performance) arrangements, including in value-for-money studies.

Source: <http://www.hm-treasury.gov.uk>

⁴ Young, Richard D. (2005). *An Overview: Oregon Shines II and Oregon Benchmarks*, Unpublished manuscript, South Carolina, Institute for Public Service and Policy Research, University of South Carolina. Accessed from: <http://www.ipspr.sc.edu/scip/publications/Final%20Oregon%20Shines%20II%20and%20Oregon%20Benchmarks.pdf>

Lesson 3: Not all activities are equally measurable

It is well known that the organizations and units within government serve highly different functions and therefore, when it comes to establishing measurement systems for organizations, some are easier to measure the outputs and outcomes of than others. To illustrate this point, Wilson’s 1989 typology of organizations is useful (Table 2). Among organizations/units, production organizations are the easiest to measure the outcomes and outputs of, for example in tax collection, sanitation, vehicle registration etc.... In such cases measuring and managing for results is easier to do than in coping organizations, such as diplomacy or research, where the outputs and the outcomes of the organization are difficult to measure. A good example of this is the US National Science Foundation (NSF), which funds cutting edge research. It is not possible to measure the impact of the organization meaningfully in the short run, rather only measures of outputs (e.g. number of papers written etc..) are available. A creative solution to this problem was the establishment of a panel of experts who examine the quality of the research portfolio on an annual basis. A more comprehensive, retroactive examination of the impact of the research portfolio is done on a 5 or 10 year basis. This suggests that the measurements and definitions of outputs/outcome should be tailored to the different profiles of government entities and interpreted with caution in the areas where such measurement is very difficult and imprecise.

Table 2: Wilsons (1989) typology of organizations⁵

		Outcomes Observable	
		<i>Yes</i>	<i>No</i>
Outputs Observable	<i>Yes</i>	Production Organizations. Examples: mail services, tax collection, sanitation, vehicle registration, revenue collection	Procedural Organizations. Examples: mental health, counselling, military (peacetime), youth penitentiary
	<i>No</i>	Craft Organizations. Examples: Field inspections, military (wartime), doctors, forest rangers	Coping Organizations. Examples: Diplomacy, Intelligence, research

Lesson 4: Simple, less complex measurement and management systems with a single entity coordinating the process reduces transaction costs and improves functionality.

Many countries suffer from overlapping mandates between actors/tools in an RBM system due to the nature of the way government is organized and the power relations between layers of

⁵ Wilson (1989) as described in Wouter Van Dooren et al, *Performance Management in the Public Sector*, (2010).

government.⁶ This creates an overwhelming demand on ministries/units for reporting, monitoring, evaluation both upward and outward. Over time, actors at the center of government have realized that coordination is critical and that a clear understanding of objectives is necessary to create a well-functioning system.

In the US state of Oregon, a Progress Board was established to oversee the implementation of the Oregon Shines II and Oregon Benchmarks. The Board provided governance and oversight for an integrated system. The 12-member board, chaired by the governor, was assisted by a staff of professionals who gather data, monitor progress, and prepare reports. The Board was tasked with clearly identifying the aims and audience for the establishment of an indicator system.

In the UK, the establishment of the PMDU's in 2001 office helped to bring focus to the various performance management initiatives underway and to supplement and draw upon existing tools (audit, evaluation etc..) for use in reporting to the Prime Minister.

The US state of Maryland's experience with StateStat – a performance measurement and management system – also underscores the importance of establishing a key actor at the center to oversee and manage the overall system. In use since 2007, StateStat was originally modeled after Baltimore's CitiStat as a way of capturing and monitoring the progress of government service delivery using frequently updated data. Through a process of continual re-evaluation by the Governor, his executive staff, and agency leadership during bi-weekly meetings, new and improved strategies emerge for delivering key public services effectively and efficiently. Located within the Governor's offices in Annapolis, a lean staff provides the logistical and analytical support for the operation.⁷

According to a report on insights from international practice on performance reporting, the six attributes of a good system for outcome and output reporting are:

1. When developing performance measurement systems, use a consistent, comparable, and structured approach to performance information reporting across all agencies and programs.
2. Include a good performance story to accompany the indicators.
3. Specify outcome indicators, and fully explain the results reported against the indicator.
4. Provide both target and baseline data.
5. Ensure effective use of technology in presenting the performance data collected.
6. Present agency performance information which includes output and activity indicators in addition to outcome indicators.⁸

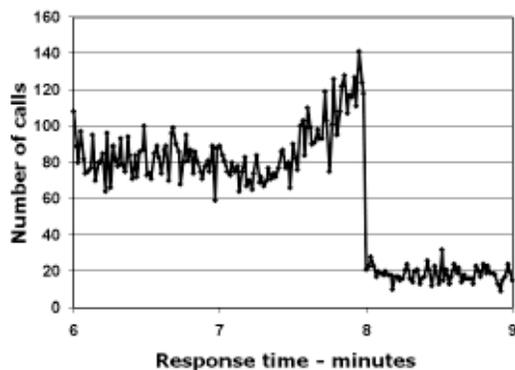
⁶ See the GET Note: Japanese Public Sector Assessment Processes “Recently Asked Questions” Series August 2010, World Bank for a description of Japan's experience.

⁷ GET Brief, “Ten Observations on StateStat,” December 2009, World Bank.

⁸ Richard Boyle, “Performance Reporting: Insights from International Practice,” IBM Center for Business of Government (2009).

Lesson 5: Proceed with caution when directly linking measurement systems to incentives

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



Getting the measurements of outputs and outcomes correct is difficult and may take years to refine. Because you will get what you measure, often with unintended consequences, countries have been reluctant to tightly incentivize measures. This is why many countries have been reluctant to tightly link performance information with monetary incentives (particularly in the case of budgeting). Moreover, there are many good examples of how measuring the wrong thing can actually counteract real improvement, for example the case in Thailand of the indicator on the number of patients seen per day by the dentist. This quickly resulted in a substantial decrease in the time spent per person and an increase in repeat visits. If this were linked directly to budgetary allocation, it could

have had disastrous spill-over effects. There is another good example of ‘gaming’ from the UK in which a target was developed for response times of ambulances to life-threatening emergency calls. In order to incentivize a quicker response time the target was set at 8 minutes or less. Under pressure to record the ‘right’ answer, data was manipulated to suggest improvements – as Figure 4 suggests, calls were reassigned to have response times under 8 minutes.⁹ As a result of such cases, most countries use performance information alongside a number of other factors when making decisions.

Conclusion

A system’s view of results based management reveals a number of different approaches and techniques used across the public sector to improve results. However, many of these approaches and techniques are adopted wholesale from other contexts, whether the private sector, or other governments, without stopping to think about how the incentive system is aligned or what the reporting burden might be on the front line – ending with increased costs and more data, but no change in performance (or for the worse!). Countries would do well to be more cautious in introducing a results based management system and keep the lessons this note reveals in mind as implementation progresses.

⁹ Bevan, G. and R. Hamblin (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.