Challenges for goal-based learning in public investments:

A behavioural perspective on performance information use

31 March 2017

Donald P. MOYNIHAN
Esptein Professor of Public Affairs, La Follette School of Public Affairs, University of Wisconsin-Madison
Background information

This paper was prepared as a background document to the OECD-European Commission Seminar on “Performance indicators” held on 31 March 2017 at the OECD Headquarters in Paris, France. It sets a basis for reflection and discussion.

About the Project

This seminar is part of a five-part seminar series in the context of an EC-OECD project “Designing better economic development policies for regions and cities”. Other sessions in the series addressed the use of: contracts for flexibility/adaptability, performance indicators, financial instruments, and insights from behavioural science. The outcome of the seminars supports the work of the Regional Development Policy Committee and its mandate to promote the design and implementation of policies that are adapted to the relevant territorial scales or geographies, and that focus on the main factors that sustain the competitive advantages of regions and cities. The seminars also support the Directorate-General for Regional and Urban Policy (DG REGIO) of the European Commission in the preparation of the impact assessment for the post-2020 legislative proposals and to support broader discussion with stakeholders on the future direction of the delivery mechanisms of regional policy.

Follow us on Twitter: OECD SMEs, Regions, Cities (@OECD_local)
Acknowledgements

My thanks to all who provided feedback on various drafts of this report, including Ivor Beazley, Richard Broos, Ronnie Downes, Peter Heil, Oliver James, Moira Lenox, Rodney Scott, Lisa Von Trapp, Lars Tummers and all of the participants at the OECD seminar “Incentivizing Performance in Public Investment Delivered at the National and Sub-National Levels” which took place in Paris on 31 March 2017.
# Table of contents

Executive summary .......................................................................................................................... 5  
Introduction: Managing across temporal and institutional horizons ........................................... 8  
Challenges for goal-based learning .............................................................................................. 8  
The new science of performance information use ........................................................................ 12  
  - Behavioural science and public policy .................................................................................. 12  
  - Behavioural public administration ......................................................................................... 13  
  - Numeric literacy ..................................................................................................................... 14  
  - The power of comparison ...................................................................................................... 15  
  - Sense of autonomy .................................................................................................................. 16  
  - Anti-public sector bias .......................................................................................................... 17  
  - Distrust of government data .................................................................................................. 17  
  - Motivated reasoning .............................................................................................................. 18  
  - Negativity bias ...................................................................................................................... 19  
Designing routines of goal-based learning ................................................................................. 20  
  - United States: Cross-agency priority goals ......................................................................... 21  
  - New Zealand ........................................................................................................................ 23  
  - Lessons for managing cross-cutting goals .......................................................................... 26  
Conclusion: Managing both cognition and design .................................................................... 27  
Notes ........................................................................................................................................ 28  
References .................................................................................................................................. 29  

Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
</table>
| Figure 1. | How New Zealand presented progress on the Results ......................................................... 25  
| Figure 2. | Short, personalised case studies to illustrate New Zealand’s Results ................................. 26  |

Boxes

<table>
<thead>
<tr>
<th>Box</th>
<th>Description</th>
</tr>
</thead>
</table>
| Box 1. | Juncker 2014 priorities ......................................................................................................... 9  
| Box 2. | Ten challenges for performance budgeting ......................................................................... 10  
| Box 3. | Particular challenges for goal-based learning for public investments .................................. 11  
| Box 4. | Insights on how people process performance data .................................................................. 14  
| Box 5. | Guidance to agencies for quarterly reviews .......................................................................... 21  
| Box 6. | Principles of well-run learning forums .................................................................................. 23  
| Box 7. | New Zealand’s ten Results ....................................................................................................... 23  |
Executive summary

Public investments involve large-scale resource allocations that are expected to provide a useful return to society. One principle of a successful public investment strategy is to facilitate goal-based learning to know if investments are being well run and delivering the expected results. Such learning depends upon the use of performance data, but the record of performance information use in the public sector has unearthed more challenges than proven successes.

The most salient challenges are the transaction costs that come from performance systems. The burdens of collecting, verifying and disseminating data become harder to justify given limited evidence of the actual use of performance data, especially among the public, or even among policy makers wary of information overload. Another set of challenges relates to temporality and problems of attribution. It is difficult to determine how resources or actions made today affect outcomes in the future. Co-ordinating multiple actors around a single performance goal is a common problem for performance systems generally. This problem is especially pronounced for public investment decisions, which by their nature involve networks of actors from multiple levels of government.

Such challenges are the reality of performance systems, stubbornly immune to simple fixes. The stock response to these challenges has been a supply-side approach – improve the quality and accessibility of performance data – or to attach financial incentives to metrics. This report explores alternative approaches, asking how performance information use is shaped by cognition – which affects what goes on inside our heads – and design features – that influence what goes on around us.

New behavioural science research on how individuals cognitively process performance data offers novel insights that suggest practical implications for how to make use of performance systems.

- **Numeric literacy**: While people say they prefer hard data, they are in fact more likely to be influenced by anecdotes or stories when judging public sector performance. A practical implication is that the presentation of data about public investments becomes more persuasive when it forms part of a broader narrative, or uses illustrative stories. Too much data results in less attention.

- **The power of comparison**: Comparative data have been shown to help make data more useful and understandable, as they provide a sense of progress and a reference point for good and bad performance. Comparable cross-country or cross-regional public investment performance metrics can motivate public and governmental interest. Performance targets raise expectations, creating a negative response even if objectively unrealistic targets are missed.

- **A sense of autonomy**: Managers who believe they have greater flexibility to influence public outcomes are more likely to make use of performance data.

- **Anti-public sector bias**: Citizens assume government services are less efficient.
• **Distrust of government data:** Citizens have lower trust in self-reported government data, partly because of distrust in government. Governments should therefore use trusted non-public sources to collect, share and evaluate data on public investments.

• **Motivated reasoning:** Our political beliefs and partisan identities affect how we process performance data. Governments should look for approaches to frame and disseminate data in ways that are non-ideological, build advanced buy-in for data across political lines and debias motivated reasoning in learning forums.

• **Negativity bias:** Citizens, managers and policy makers respond more to numbers showing bad performance than to evidence of positive performance. Governments should direct resources to problem-solving areas that performance data suggest pose the greatest risks.

In addition to better building performance systems to acknowledge behavioural biases, governments can also design an environment conducive to learning from cross-cutting goals like public investments. The experiences of two very different countries – New Zealand and the United States – suggest common lessons for managing cross-cutting goals:

• **Focus on a manageable number of goals:** In the United States, just 16 cross-agency priority goals were selected for the 2015 budget. The focus in New Zealand was even tighter, with just ten goals over five years. This emphasis on a small number of goals recognises that leadership’s attention is limited and an excess of goals actually reduces the ability to make progress.

• **Participatory selection of goals:** In both cases, goals were selected based on an inter-agency dialogue in order to increase agency commitment.

• **Regular learning forums:** The United States had formal quarterly reviews to track progress. While there was no legal requirement for updates in New Zealand, inter-agency groups at multiple levels met regularly to discuss progress and strategise about next steps, and a group of about 100 made up a group called “the Results Network” that prepared progress reports every six months.

• **Public commitment to targets, with regular updates on both implementation plans and metrics:** Metrics were publicly available in both countries. In the case of New Zealand, the goals were written primarily for a public audience, and were visible enough to capture interest from the media, making them difficult to abandon.

• **Named goal leaders:** Both New Zealand and the United States experimented with having a named leader responsible for goals. In practice, this leadership was often shared. In the United States, White House officials shared responsibility with a senior manager of a key agency. New Zealand started with individual leaders, but evolved to a model where a group of leaders were given a collective responsibility. The gradual shift in both cases suggests that the accountability benefits of a single leader might be outweighed by the collaborative capacities that come from a team of leaders.

• **Mixture of true outcomes, intermediate outcomes and processes:** In both countries outcomes were accompanied by cross-cutting management goals that aimed to improve processes or intermediate outcomes that had a direct relationship with public outcomes. Such measures allow a sense of momentum since progress can be made in the short to medium term, and they reduce the temporal problem of public investments.
- **Targets to motivate progress**: The United States focused on rates of progress rather than specific numerical targets, usually relying on narrative updates. New Zealand set more explicit targets, but avoided the bias of allowing a near-miss to be confused with failure. “In New Zealand, progress tended to be described relative to the baseline rather than the target. A huge improvement that just fell short of the target was a reason for celebration rather than punishment” (Scott and Boyd, 2017: 9).
Introduction: Managing across temporal and institutional horizons

Public investments are generally considered to be “capital expenditure on physical infrastructure (e.g. roads, government buildings, etc.) and soft infrastructure (e.g. human capital development, innovation support, research and development, etc.) with a productive use that extends beyond a year” (OECD, 2014b: 2). The topic of public investments is large, and this report does not seek to undertake either a comprehensive review nor to consider the many ways performance data may facilitate public investments. Instead, the focus is to offer evidence from new research that speaks to one goal of public investments: to facilitate goal-based learning as a strategy to improve the use of public resources. For example, the 2014 OECD Recommendation of the Council on Effective Public Investment across Levels of Government identifies as a best practice that countries:

… clarify the outcomes to be achieved through public investment and pursue mechanisms to achieve them. Those mechanisms can include results-oriented investment strategies with clearly defined policy goals, well-designed tendering procedures, effective monitoring systems, high-quality ex post evaluation, regular reflection on and upgrading of investment choices, active exchange of information and on-going and mutual learning among actors involved in public investment. (OECD, 2014a: 11)

Indeed, the OECD identified one principle of action for effective public investments to be: “Focus on results and promote learning from experience across levels of government” (OECD, 2014b: 22).

There is, therefore, strong hope in the possibility of performance data as the key to improving public investment outcomes, whether it be through performance budgeting, assessing project implementation or to broadly facilitate learning. Much depends on the assumption of the public that policy makers are able to make sense of performance data. The next section maps out some of the challenges to goal-based learning that have been observed from the experience of implementing performance systems in government. Public officials have wrestled with these challenges largely by trying to build better reporting frameworks or altering incentives. However, a behavioural perspective is a source of new insights about how people cognitively process information on public sector outcomes. Behavioural science is a broad and growing field that incorporates psychology, economics, neuroscience, public policy and law, among others. This report focuses on the intersection of behavioural science and public management, sometimes called behavioural public administration (Grimmelikhuijsen et al., 2017; James, Jilke and Van Ryzin, 2017) and only on research related to how performance data are used. While the focus is therefore narrow, it allows the most in-depth single assessment to date on how citizens and policy makers process performance data.

The focus of behavioural science on cognition – what goes on inside our heads – is balanced by the last section of the report, which examines efforts to design broader environments conducive to goal-based learning. The efforts of New Zealand and the United States to encourage attention to high-level cross-cutting goals suggests some principles of design that are relevant to public investments.

Challenges for goal-based learning

This section offers an overview of efforts to generate goal-based learning in government, which often take the form of performance budgeting initiatives. The central lesson is that even though there are substantial challenges to performance budgeting, especially for public investments, they still remain popular. Most countries have some sort of performance
framework in place that seeks to clarify goals and measures, often with some form of key national indicators.

The European Union is similar to many OECD countries in that it has a broad performance framework. The combination of Europe 2020 framework and the Juncker 2014 priorities provides a broad set of goals for the European Commission, which recently launched the 2014 “EU budget focused on results”. The EU budget is of special interest because it is primarily centered on public investments (OECD, 2016a: 2), with over half of the budget allocated to European Structural and Investment (ESI) funds.

<table>
<thead>
<tr>
<th>Box 1. Juncker 2014 priorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. A connected digital single market.</td>
</tr>
<tr>
<td>3. A resilient energy union with a forward-looking climate change policy.</td>
</tr>
<tr>
<td>4. A deeper and fairer internal market with a strengthened industrial base.</td>
</tr>
<tr>
<td>5. A deeper and fairer Economic and Monetary Union.</td>
</tr>
<tr>
<td>6. A reasonable and balanced free-trade agreement with the United States.</td>
</tr>
<tr>
<td>7. An area of justice and fundamental rights based on mutual trust.</td>
</tr>
<tr>
<td>8. Towards a new policy on migration.</td>
</tr>
<tr>
<td>9. Europe as a stronger global actor.</td>
</tr>
<tr>
<td>10. A Union of democratic change.</td>
</tr>
</tbody>
</table>


A basic goal of performance systems is to identify what strategic goals public resources should be invested in, offering a clear set of goals that stakeholders can work towards (Moynihan and Beazley, 2016; OECD, 2014b). Moynihan and Beazley assess the performance budgeting practices of seven different countries, identifying a series of common problems, summarised in Box 2.
Box 2. Ten challenges for performance budgeting

1. Setting objectives: Performance budgeting reforms are driven by a variety of motivations and, depending on the political and administrative context, may evoke different responses. For countries with strong centralised governments, the primary objective is typically to make the budget more responsive to national goals and policy priorities. Other countries see it as giving managers more flexibility to allocate resources and innovate to improve service delivery. A primary challenge for governments is to decide explicitly what they want performance budgeting to achieve.

2. Looking beyond the budget: The annual budget process is too compressed to allow for considered assessment of programme performance. Managers are the most likely users of performance information; zeroing in on the budget process understates the value of performance budgeting for programme management.

3. Capacity constraints: Governments often underestimate how much administrative and analytical capacity is necessary to operate a performance budgeting system successfully. Even OECD countries rarely set aside enough resources, and many countries may lack the capacity for the type of ambitious approach they adopt. Not only should more attention be paid to capacity, but performance budgeting reform may also need to be simplified.

4. Information overload: Countries typically produce too many metrics, leading to information overload. The complexity of programmes and indicators must be tightly controlled.

5. Prioritising strategic goals: Different performance data serve different needs; metrics for the most salient service delivery areas are more likely to be used.

6. Managing performance perversity: Too much emphasis on performance can motivate gaming behaviour. Manipulation of performance data is rare, but when it comes to light it can create severe trust and legitimacy problems for governments.

7. Routinising performance information use: Building separate programme evaluation, audit and learning routines can address policy and management questions and provide valuable complements to performance budgeting.

8. Changing behaviour: Changing public employee attitudes is difficult. It takes persistent efforts, often over a decade or more to introduce a performance-oriented culture, supported by general public service reforms. The pace and ambition of performance budgeting changes should reflect this reality.

9. Balancing political and bureaucratic support: A political champion can help direct attention to performance budgeting – but it can also create opposition and abandonment if reform is viewed as a partisan tool.

10. Learning from experience: The countries that have progressed the most have done so by identifying and resolving the shortcomings of previous reform efforts, rather than repeating past failures or starting anew.


It is worth considering how some of these challenges combine to become especially salient for the special category of public investments, though it is worth noting that little systematic research has been done on performance efforts in this area. These particular challenges are summarised in Box 3.
Box 3. Particular challenges for goal-based learning for public investments

- Temporal problem: Outcomes may take years to materialise, even as policy makers operate in one-year frameworks.
- Problem of attribution: The causal connection between an investment and outcome is uncertain.
- Multi-actor co-ordination: High-level outcomes require multiple levels of government, different areas of government and the private sector, to share information and co-ordinate effort.
- Transaction costs: The significant administrative costs of collecting, auditing and disseminating performance data may not appear to be justified by use.

It is difficult to connect the effect of a public policy to an outcome when the effect may take years to materialise and may be crowded out by other factors. This **temporal problem** is juxtaposed with institutional realities where policy makers must operate under a more constrained time frame, taking decisions through annual budgets and looking to the next election. One solution is to monitor outcomes over a multi-year period and use a multi-year budgeting framework. Even so, the temporal problem creates a potential for public investment trade-offs between spending items that have more immediate and tangible products (such as roads or transport) and those where it is difficult to discern the outcome in the short run (such as investments in human capital).

The **problem of attribution** is that it is difficult to empirically connect how public programmes lead to outcomes. Performance data by itself do little to provide insight into whether a programme made a difference – they can only show if measures rose or fell. Ideally, programme evaluations should play such a role, along with performance data that offer a comprehensive understanding of both how and why performance has changed over time. Managers are less likely to make use of data when they cannot clearly connect how a programme matters to outcome (Moynihan and Kroll, 2016). This problem is true for any moderately complex public outcome, but is more severe for public investments where the causal link between the investment and intended effect may be unproven, where broader societal factors also shape the outcome, and especially when the outcome is the result of a complex interplay of multiple factors, e.g. student educational performance.

High-level public outcomes typically require **multi-actor co-ordination**, e.g. different levels of government, or different units within government, and private partners. This is certainly true of public investments, where 72% of investments across OECD countries are made at the subnational level: “Cross-sectoral, cross-jurisdictional and intergovernmental co-ordination is necessary, but difficult in practice” (OECD, 2014b). This problem is compounded when lower levels of government have limited capacity. The use of performance targets may sometimes worsen this problem, reinforcing jurisdictional boundaries as agencies focus attention on tasks they are being directly held responsible for and shy away from co-operative efforts. Creating high-level cross-cutting goals may reduce this problem, but only if the actors involved understand they are contributing toward that goal. For example, the EU lacks mechanisms to ensure co-ordination between ESI Funds and national performance budgeting frameworks (OECD, 2016a). See the section on “Designing routines of goal-based learning” for more on cross-cutting goals.
The transaction costs of performance systems are frequently underestimated. It requires time and administrative capacity to collect data, validate their accuracy and disseminate them in a useful fashion. A perennial risk is that administrative actors spend more time collecting the data than using it.

Such complaints have been leveled at regional development policies (OECD, 2009: 54), and against the new “EU budget focused on results”: “In general terms, many stakeholders have expressed the need for a radical simplification and streamlining of the EU system of budgeting for results, allowing for a sharper and more objective focus on strategic goals, achievements and results, leading to better and more informed decision-making, clearer communication, and engendering greater trust in the effectiveness of the system” (OECD, 2016a: 3). Any performance system offers the possibility of measurement creep, and some countries have sought to reduce the number of key indicators they focus on (Moynihan and Beazley, 2016). These problems may be more prominent in the multi-actor setting of public investments in the EU, since much of the costs are distributed to individual governments that find themselves asked to report back on too many indicators they do not see as useful (OECD, 2014b: 22).

Governments typically take on the challenges of performance budgeting using what might be characterised as a supply-side approach: more and better data that are easily accessible and better logic chains between inputs and outcomes. A supply-side approach is not necessarily wrong: you can’t have good performance systems without good data. But it assumes that the use of performance data for learning is driven largely by the nature of the data itself. Creating more and more metrics while neglecting the human factor means that efforts to improve performance systems often increase transaction costs without necessarily resulting in more use of performance data (Moynihan and Lavertu, 2012).

In some cases, a supply-side approach has been supplemented as an appeal to extrinsic motivation, through the use of financial incentives tied to budgets. For example, 6% of ESI Fund allocations are placed in a “performance reserve” and allocated in the case of good performance, or reallocated to other areas in the case of poor performance. Since the “performance reserve” technique is new, its effectiveness is unproven thus far (OECD, 2016a), but in some instances, these extrinsic incentives have led to gaming behaviour (Heinrich and Marschke, 2010).

The mixed, at best, record of performance systems suggests that any effort to pursue goal-based learning that does not move beyond a supply-side or incentive-based perspective has inherent limitations. The next section seeks to expand the understanding of how performance information use occurs, incorporating a behavioural science perspective. Such a perspective offers another window of the complexity of performance systems. While the tenets of a supply-side or incentive-based approach are relatively straightforward, it is less clear how to build a demand for data, or how to correct the types of biases described in the sections that follow.

The new science of performance information use

Behavioural science and public policy

Behavioural science draws from psychology, economics and other fields to focus on how individuals construe the world and trace how this affects the decisions they take. The unit of analysis is therefore the individual, with a strong emphasis on cognitive processes. When it comes to decision making, behavioural science starts with the premise that people
are not fully rational, if rationality implies making a full and exhaustive assessment of costs and benefits of each option to optimise a preferred outcome (Simon, 1947). Instead, people use heuristics to manage such complexity. Sometimes these heuristics serve as useful shortcuts to efficient decisions, but they can also cause cognitive biases that lead to systematic errors (Tversky and Kahneman, 1975).

Some of the insights of behavioural science fit well with previously identified issues with public investments. For example, the temporal problem of investments may be related to a basic tendency to discount the future relative to the present, a problem that worsens under conditions of scarcity (Shah, Mullainathan and Shafir, 2012). The problem of attribution may be related to what the OECD has pointed to as “optimism bias” that causes decision makers to uncritically assess the returns on potential investments (OECD, 2014b: 12).

Other well-known biases in decision making are a preference for the status quo and a tendency to seek information to confirm prior beliefs (Kahneman, Knetsch and Thaler, 1991). For example, experiments have shown that when individuals are given a range of investment portfolios, they tend to favour the one labeled as the status quo (Samuelson and Zeckhauser, 1988). This may be because people develop an emotional attachment to prior decisions. It may also be that decision makers do not want to publicly reverse themselves. A natural tendency for policy makers is therefore to stick with public investment decisions that have been previously taken, even if they are not working out well. They will justify these decisions through what is called conformation bias – relying on information that supports prior decisions while discounting critical information. For example, such behaviours have been found in how public officials treat investments in technology (Keil, Mann and Rai, 2000; Moynihan and Lavertu, 2012).

**Behavioural public administration**

Some governments have set up special units to take advantage of behavioural science. But most of the initiatives that emerge have been about the design or implementation of policy, not with management processes such as performance budgeting (OECD, 2017b). However, a newer area of research, sometimes called behavioural public administration, has started to apply behavioral insights to the workings of public organisations (Grimmelikhuijsen et al., 2017; James, Jilke and Van Ryzin, 2017). A sub-set of that literature specifically address questions relevant to goal-based learning, examining the cognitive processes related to performance information use.

This research is distinct from an existing literature on performance information use in its use of experimental designs to understand the psychology between individual behaviour in public organisations. It seeks to get inside the head of those using the data. As a result, most of the work examines how people process different types of data, presented in different ways.

Because public investments include high-level outcomes, it is as important to look at how citizens, as well as policy makers, make sense of data. Indeed, a best practice of effective public investment is to engage with stakeholders (OECD, 2014b), and performance data can play a role in doing so. But a central goal of a behavioural approach is to identify cognitive processes that are widely generalisable from one setting to another. The cognitive process of a citizen and a policy maker should generally be assumed to be similar. It is the case that policy makers have access to more information and the requirement to negotiate with other policy that may help to check their biases. However, most evidence suggests that even well-informed policy makers are similar to members of
the public in how they process data (Sheffer et al., 2017). Indeed, expertise can actually worsen biases, since it results in overconfidence (Tetlock, 2005). These sections consider what factors make it more or less likely that people use data, but also provide insight into how they use data, summarised in Box 4.

**Box 4. Insights on how people process performance data**

- **Numeric literacy**: While people say they prefer hard data they are in fact more likely to be influenced by anecdotes or stories when judging public sector performance. A practical implication is that the presentation of data about public investments becomes more persuasive when it forms part of a broader narrative, or uses illustrative stories. Too much data result in less attention.

- **The power of comparison**: Comparative data have been shown to help make data more useful and understandable, as they provide a sense of progress and a reference point for good and bad performance. Comparable cross-country or cross-regional public investment performance metrics can motivate public and governmental interest. Performance targets raise expectations, creating a negative response even if objectively unrealistic targets are missed.

- **A sense of autonomy**: Managers who believe they have greater flexibility to influence public outcomes are more likely to make use of performance data.

- **Anti-public sector bias**: Citizens assume government services are less efficient.

- **Distrust of government data**: Citizens have lower trust in self-reported government data, partly because of distrust in government. Governments should therefore use trusted non-public sources to collect, share and evaluate data on public investments.

- **Motivated reasoning**: Our political beliefs and partisan identities affect how we process performance data. Governments should look for approaches to frame and disseminate data in ways that are non-ideological, build advanced buy-in for data across political lines, and debias motivated reasoning in learning forums.

- **Negativity bias**: Citizens, managers and policy makers respond more to numbers showing bad performance than to evidence of positive performance. Governments should direct resources to problem-solving areas that performance data suggest pose the greatest risks.

**Numeric literacy**

People, even experts, are not naturally number crunchers and have difficulty estimating basic statistics (Hoffrge et al., 2000). We become overwhelmed when we face too many metrics, creating the familiar problem of information overload. When asked if they prefer hard data or an anecdote to take a decision, people call for the former but are actually more influenced by the latter. Surveys find that citizens state that they prefer quantitative over qualitative information to evaluate performance, but experiments show that anecdotes or stories generate a greater emotional response, are more memorable and are more influential in the performance assessments of public organisations (Olsen, 2016).

Citizens also tend to find round numbers more salient in a way that alters their decisions. This is sometimes referred to as left-most digit bias. For example, an experiment that assessed public support for a hypothetical public investment stimulus programme found that support tended to move up in a discontinuous fashion, with opposition jumping more
when the cost of the stimulus went above a round number (Malhotra and Margalit, 2010). In evaluating school performance, Olsen (2013) shows that citizens exhibit a similar pattern of left-most digit bias. Substantive shifts in grade scores that did not cross a threshold (e.g. if grades moved from 5.1 to 5.9) went unnoticed in citizen evaluations, but when grades moved above a threshold (e.g. increased from 5.9 to 6), citizens became substantially more positive in their assessments.

**Practical implications:** Governments need to be realistic about the capacities of citizens to manage data. An excess of data will lead to less attention than a smaller amount of data that does not tax the cognitive load of the user. More can be expected of policy makers and their staff, but their complaints about information overload reflect that busy decision makers may lack the time to process data, even if they have higher numeric literacy (Moynihan and Beazley, 2016). Some governments have responded by reducing the number of metrics they offer policy makers (Moynihan and Beazley, 2016).

When asked to make simple statistical estimates, people do better when they are presented with frequencies (e.g. 60 out of 100 people will be eligible to receive a programme, but only 30 of those actually receive the services) rather than as probabilities or percentages (Hoffrage et al., 2000). Numbers by themselves are less powerful than narrative, so when it comes to presenting data, metrics should be tied to illustrative stories or anecdotes.

**The power of comparison**

Presenting data in different formats can make it more or less interesting, useful and motivating. A list of numbers by itself is of limited value, since an individual needs some sense of context to make sense of the data. Comparisons – with peers, with past performance or with some future target – provide such a context. Comparative data have been shown to help make data more understandable, since they provide a sense of progress and some reference point for what is good or bad performance (Nielsen, 2014).

Citizens view comparative performance data as more persuasive (James and Moseley, 2014; Olsen, 2017). But not all comparisons are created equal and comparisons with peers appear to have a markedly stronger impact on citizen evaluations of programme performance than do comparisons with an organisation’s own past (Charbonneau and Van Ryzin, 2015; Olsen, 2017). This seems to be because people are inherently more interested in how they stack up against peers, especially salient ones like neighbouring or similar governments, or historical rivals. For example, Danes looking at comparisons of public sector performance find their neighbour Sweden to be more useful than other countries (Hansen, Olsen and Bech, 2015).

Anecdotally, there is already evidence of the power of cross-country comparisons. Perhaps the best example is the Programme for International Student Assessment (PISA), a cross-national educational rankings of student performance that has had a meaningful influence on education policy management. PISA rankings gain a good deal of attention as national media covers how well a particular country stacks up against other countries, especially if national performance is perceived as falling below expectations (Martens and Niemann, 2013). PISA rankings are potent in policy-making debates partly because they are easy to understand, and deal with a politically salient function.

Comparative data also seem to motivate public managers, even if the data are not directly tied to rewards. In one experiment, school principals were offered the opportunity to download performance data about their school on a website (Andersen and Moynihan, 2016). One group received performance data only about their school, while another group
was also offered comparative data showing how their school ranked compared to others. The rate of downloads was monitored, unbeknownst to the subjects. Those offered comparative data tended to download more performance data, including other types of performance data than those who were just offered data about their own school.

Performance targets provide another form of comparison, since they establish aspirational goals that will later be used to compare against outcomes. One experiment showed that the same actual level of performance may be perceived as more or less impressive if accompanied by performance targets that are respectively modest or ambitious (Moynihan, 2015). One group of subjects was given data showing consistently improving performance for a programme. The treatment group received the same performance data, but also were shown targets that made clear that even though the underlying performance was improving over time, it failed to match targets. As a result, the treatment group was less willing to allocate resources to the programme.

Targets are not just quantitative metrics. They can also be broad categories indicating higher or lower achievement. A programme might be given green, yellow or red lights to communicate progress, or labeled as succeeding or failing. For example, in the United States, broad categorisations that identified schools as failing under federal rules were rendered close to arbitrary given the different ways states calculated cut-off points. But the labels still mattered to citizens. Schools labeled as failing were less able to raise resources via referenda (Kogan, Lavertu and Peskowitz, 2016) and saw greater exit of students (Holbein, 2016). Even administrators are prone to place too much weight on such categories. One study of Texas schools found that schools that were essentially equivalent were treated differently when it came to resource allocation decisions taken by administrators, depending on whether they fell just above or below a step in the performance rating system (Craig, Imberman and Purdue, 2015). The combination of target categories and limited numeric literacy described above can substantively affect citizens’ and policy makers’ perceptions, even if differences in the categories are not substantively meaningful.

Practical implications: Governments can draw greater attention to performance data by presenting it in comparative perspective, especially using peer comparisons. About one-third of OECD countries use some form of national indicators, and the majority of these are internationally comparable, including many that are aligned with the Europe 2020 objectives (OECD, 2016b; 2017a). One risk of peer comparisons is that countries or regions may start from very different places, so historical comparisons provide a good balance. The use of targets may create a risk for programmes, for when those targets are not met – even if performance scores are improving – this reduces support for a programme among subjects. If such a risk is apparent to practitioners, they may wish to set incremental and realistic targets.

Sense of autonomy

A consistent claim from the performance management literature is that public managers who feel like they have more discretion also say they make greater use of performance data. There are a couple of psychological reasons why more autonomy should facilitate greater use of performance data. Most obviously, managers who feel that they are powerless to change outcomes have little motivation to learn more about how well their organisation is doing, and how to improve it. Political scientists argue that the provision of autonomy encourages public managers to invest in building their expertise, confident that they can actually use it to influence public outcomes (Gailmard and Patty, 2012). Public managers
may also simply want to understand if the decisions they have taken are related to positive or negative outcomes, reflecting a trial-and-error approach to learning.

Until recently, however, the evidence for the relationship between autonomy and performance information use came from survey-based studies that were unable to discern if this relationship is correlational or causal or in which direction causality might run. To solve this problem Andersen and Moynihan (2016) developed a field experiment to test if giving managers more autonomy resulted in their using performance data more. School principals were given resources to hire a new employee and randomly assigned varying levels of autonomy in the type of employee they could hire. For example, some had to hire new employees with traditional educational qualifications, while some could hire new employees with any educational background. The study found that principals given more autonomy were more likely to download performance data about student well-being in their schools even months later.

**Practical implications:** Performance management reforms are often presented in terms of holding managers accountable for performance, but in return they are given more flexibility from administrative rules and procedures. Too often governments fail to keep this bargain. In practice, performance systems are often imposed in addition to, rather than instead of, traditional compliance requirements (Jakobsen and Mortensen, 2015; Moynihan, 2008). Finding ways to offer managers more autonomy restores the promise of performance management.

**Anti-public sector bias**

A number of studies from Denmark and the United States have shown that individuals make more negative assessments of organisations if they are presented as public rather than private, even if they are given the same performance data (Marvel, 2016; Hvidman and Andersen, 2016). This has been dubbed an implicit anti-public sector bias: “Citizens automatically and unconsciously associate public sector organizations with inefficiency, inflexibility and other pejoratives, and these automatic associations color their assessments of public sector performance” (Marvel, 2016: 143).

**Practical implications:** It’s unclear what the practical implications are here, beyond the need to be aware of this bias. One possibility might be to emphasise the public-private partnerships that arise in public investments. On the other hand, such a strategy creates the risk of “submerging” the important contributions of public investments. In the long run, governments can better address this bias by consistently offering evidence of public sector achievements (Mettler, 2011).

**Distrust of government data**

Citizens have lower trust in self-reported government performance data. This may partly reflect an underlying suspicion of government that is described above. Distrust of government data appears to be greater as the task grows more complex. For simple and easily verified tasks such as clean streets, governments’ self-reports are viewed as being as credible as other sources (Van Ryzin and Lavena, 2013). But for more complex indicators, like citizen satisfaction, citizens trust self-reports less than reports from an independent, non-partisan source (James and Van Ryzin, 2017a). The public also becomes more suspicious of government reporting if the metrics show high levels of performance (James and Van Ryzin, 2017a).
**Practical implications:** Use trusted non-public sources or credible independent public sources to share data on public investments. Evaluations by third parties, or by the EU for investments in member countries, provide another means of building trust in public investments.

**Motivated reasoning**

Our political beliefs and partisan identifications affect how we process performance data. For example, conservatives and liberals react to the same performance data about public organisations differently. The credibility of those explaining the data depends on whether they are assumed to share the same ideological beliefs of the audience. This is highly relevant in public investment settings where there are multiple stakeholders who may hold differing beliefs about the purposes of spending.

Motivated reasoning is a specific form of confirmation bias where individuals select and process information to fit with pre-existing beliefs, most frequently partisan or ideological beliefs (see Lodge and Taber, 2000).

The term “motivated reasoning” is a shorthand term that captures the tendency for individuals to select and use information in ways that fit with ideological beliefs. For example, partisans tend to be less likely to attribute responsibility for negative outcomes to fellow party members, but are more willing to blame opposing parties (Tilley and Hobolt, 2011). In the same way, citizens presented with positive performance data about a local government are more likely to see their assessment of the incumbent party improve if they already identify with that party (James, 2011).

Political beliefs also shape how members of the public assess the performance of specific public investments. One significant investment in public health in the United States is the Affordable Care Act. James and Van Ryzin (2017b) undertook an experiment to understand how those with different political views made use of performance data to assess the policy. They found that when Democrats were given information designed to prime their political identity, they selected performance data that portrayed the healthcare law more favourably and evaluated such favourable performance data as having greater credibility.

As noted above, citizens tend to have a general anti-public sector bias, but the extent of this bias depends partly on their ideological beliefs. When given the same performance data, conservative citizens tend to more positively assess the performance of a hospital if they are told it is a private provider rather than a public one, while liberals interpret performance data more generously for public hospitals (Baekgaard and Serritzlew, 2015). A replication of this experiment for politicians shows the same results and even shows that the provision of additional performance data at odds with ideological beliefs is not enough to debias their assessments (Baekgaard et al., 2017).

An experiment with Swiss legislators found that the addition of performance data made them more willing to deviate from the status quo allocation to recommend a different level of investments in transportation infrastructure (Demaj, 2017), though not in a predictable pattern. In some cases, it caused legislators to recommend more funding, sometimes they recommended less. One reason why performance data might facilitate larger budget changes is that it helps to increase confidence in pre-existing beliefs, especially in ambiguous and complex settings, even if the data do not do too much to resolve such complexity (Sniezek, 1992; Tsai, Klayman and Hastie, 2008). For the Swiss legislators,
partisanship also played a role. Performance information had the greatest effect when it helped to provide support for easy budget trade-offs that were consistent with political preferences (e.g. investing in road infrastructure at the expense of environmental spending), helping to reinforce pre-existing partisan positions.

Motivated reasoning also shapes who policy makers find credible. Nielsen and Moynihan (2016) employed a survey experiment of local elected officials to test the role of ideology in shaping how advocacy matters. Elected officials were given performance data about local schools and some were provided a statement from a stakeholder – a union member – casting doubt on the validity of the test scores. For liberal politicians, the stakeholder was credible because they shared the same ideological space, resulting in the politicians becoming less likely to use test scores. By contrast, conservative politicians ignored the comments of the union stakeholder. The result is that advocacy about the data actually increased the distance between how liberal and conservative politicians used the data.

**Practical implications:** Stakeholders will not approach data neutrally, but will frame it to pursue their ideological goals. The credibility of arguments about the validity of data will be influenced by the political relationship between the advocate and the person hearing the argument. Simply providing people with more information will not debias motivated reasoning (Baekgaard et al., 2017). So, what can help? Governments should look to frame and disseminate data in ways that are non-ideological, such as building advanced buy-in for data and goals across political lines.

There is some evidence to support two other possible options. James and Van Ryzin (2017b) suggest that the value of policy and how they might benefit from it seems to moderate motivated reasoning. A group of citizens who were asked to think about their healthcare needs tended to have much more similar views of the US Affordable Care Act relative to citizens who were asked to think about their political identity. A related approach is to force people to think about policy implementation. Fernbach et al. (2013) find motivated reasoning is sustained by an “illusion of explanatory depth” that can be punctured if people are asked to talk through how their preferred policies would be implemented in as much detail as possible. Such discussions cause people to confront gaps in their own knowledge, making them less attached to the prior position. The challenge with such approaches is whether it is possible to structure how people receive information, or think about policies. For the public, this may be close to impossible given that people choose sources of information in ways that tend to reinforce motivated reasoning. But for policy makers, and especially public managers, it is more feasible to engage in structured dialogues about performance, which I have characterised elsewhere as learning forums (Moynihan, 2008). The final section offers additional discussion of learning forums.

**Negativity bias**

Citizens, managers and policy makers are more responsive to numbers showing bad performance than to data showing good performance. This arises from a psychological trait known as loss aversion, and in practice results in policy makers viewing performance data through a negativity bias.

Loss aversion is a bias that makes losses loom larger and more psychologically painful than equivalent gains. Loss aversion generally makes people more averse to action that might increase losses, or makes them engage in risky behaviour to minimise losses (Tversky and Kahneman 1975).
How does loss aversion matter to the use of performance data? Citizens, politicians and managers are most attentive and responsive to performance data showing poor performance. Among citizens, both survey experiments (Olsen, 2015; James and Moseley, 2014) and actual elections offer evidence of this pattern (Kogan, Lavertu and Peskowitz, 2016). For example, Olsen (2016) shows that simply presenting the same hospital data in terms of patient dissatisfaction rather than patient satisfaction reduces citizens’ evaluation of the performance of the hospital.

Citizens tend to assign less favourable ratings of political incumbents if shown performance data that imply poor performance, but generally to not provide equivalent credit if the data show positive performance (James, 2011; James and Moseley, 2014). This pattern holds in elections where incumbent UK local government officials tended to lose vote share when local governments were publicly rated as poorly performing, but did not gain equivalent votes in cases of positive ratings (Boyne et al., 2009). One way for politicians to respond to negativity bias is to shift blame for poor performance on to others. A UK experiment suggests that shifting blame to bureaucrats is rational for elected officials. Here, subjects were given evidence of poor performance in street maintenance. Relative to a control group of citizens who were given no information on the management of the programme, blame for politicians increased if subjects were told that politicians managed street maintenance, stayed the same if they were told that a private contractor managed services, and declined if told that street maintenance was run by local government employees (James et al., 2016).

Given the prevalence of negativity bias among voters it is perhaps unsurprising that elected officials not only demonstrate similar tendencies, but look to find ways to minimise blame or shift responsibility. In experiments where politicians were given actual performance data showing local schools, they focused their attention on low-performing schools (Nielsen and Baekgaard, 2015). Poorer performing schools were prioritised both for more resources and administrative reforms relative to schools that had performed better. One caution from this finding is that education is an essential public function. For policy areas seen as optional, policy makers may respond by deciding to withdraw resources. The key point is that they take actions that will minimise political blame from poor performance. Another set of survey experiments of elected officials using an equivalent design found that politicians became more willing to attribute responsibility to school principals for performance outcomes when performance was poor (Nielsen and Moynihan, 2016).

**Practical implications:** Public confidence in government is defined by public sector failures, not successes. Governments can minimise negativity bias by presenting data in terms of level of achievement (e.g. satisfaction rates) rather than failure (dissatisfaction rates). But negativity bias has more profound implications for how governments think about how to allocate resources to performance. It suggests government will receive little credit beyond a certain level of performance and will gain the most by taking a problem-solving approach to performance. A problem-solving approach suggests minimising failure rather than try to maximise performance, focusing more attention to the lower end of the performance distribution. Governments can still use data with this approach to: 1) identify areas where process or performance indicators are lagging or suggest a risk of failure; 2) assess if targeted strategic efforts to fix those problems are making a difference.

**Designing routines of goal-based learning**

The previous section identified how cognitive processes affect how performance data are processed. This section looks at other ways governments have tried to influence the use of performance data.
of performance data, specifically high-level outcome goals that cut across multiple agencies as public investments do. Two examples will be described, one from the United States and one from New Zealand. In the US case, there is evidence that new routines increased the use of performance data (Moynihan and Kroll, 2016), while New Zealand demonstrated clear progress on ten key performance indicators over the course of five years (Scott and Boyd, 2017).

United States: Cross-agency priority goals

The US federal government does not have a formal public investment office or strategy, owing partly to a political system fragmented by a separation of powers between the executive and legislative branches at the federal level, and a federalist system that distributes power to states and localities. At the same time, it has wrestled with various incarnations of its performance system for a quarter century, since the passage of the Government Performance and Results Act of 1993, which was updated by the GPRA Modernization Act of 2010 (Moynihan, 2016). Even as a new administration brings a different perspective to governmental performance, it remains bound to the statutory processes for assessing performance.

The US performance system requires that each agency complete strategic plans, measure performance on an annual basis and make these data public. The Modernization Act added new requirements, including a requirement that agencies work together on cross-agency priorities. The cross-agency priority goals cumulatively represent a government-wide strategic plan.

Cross-agency priority goals are similar to public investment goals in that they are high-level goals that require the co-ordination of multiple actors. To facilitate co-ordination, the Modernization Act requires a number of actions. Agencies must report on how they contribute to the goals (if relevant) using a standard set of metrics. While some of the goals were long-term outcomes, the Modernization Act also allowed long-term management priorities to be treated as cross-agency priority goals, a nod to the difficulties inherent in co-ordinating these goals. For example, a 2015 budget cross-agency priority goal was “Modernize the federal permitting and review process for major infrastructure projects to reduce uncertainty for project applicants, reduce the aggregate time it takes to conduct reviews and make permitting decisions by half, and produce measurably better environmental and community outcomes.” The steering committee for this goal included 13 federal agencies who in turn worked with state, local and tribal governments.

The Modernization Act also requires that cross-agency priority goals are reported and reviewed by relevant officials each quarter, and are assigned a specific goal leader. The quarterly reviews are an example of an effort to deliberately adopt a new social routine in government. Like most performance systems, previous iterations of the US approach had involved routines that created and disseminated data, but did little to build routines for use. Rather than assuming that the existence of data will lead to use, the requirement of quarterly reviews assumed that performance information use must be deliberately designed into organisational routines. By requiring regular reviews of a small number of key goals, the Modernization Act forces leaders to pay attention to these metrics and to generate feedback about progress. The US Office of Management and Budget offered some guidance on how to set up and run these reviews (Box 5).

| Box 5. Guidance to agencies for quarterly reviews |
Review with the appropriate goal leader the progress achieved during the most recent quarter, overall trend data and the likelihood of meeting the planned level of performance.

Hold goal leaders accountable for knowing whether or not their performance indicators are trending in the right direction at a reasonable speed, and if they are not, for understanding why they are not and for having a plan to accelerate progress on the goal.

Hold goal leaders accountable for knowing the quality of their data, for having a plan to improve it if necessary, and for filling critical evidence or other information gaps.

Hold goal leaders accountable for identifying effective practices by searching the literature, looking for benchmarks and analysing disaggregated data to find positive outliers across performance units.

Hold goal leaders accountable for validating promising practices with replication demonstrations or other evidence-based methods.

Review variations in performance trends across the organisation and delivery partners, identify possible reasons for the variance, and understand whether the variance points to promising practices or problems needing greater attention.

Include evaluation staff to share and review performance information and evaluation findings, to better understand performance issues that evaluation and research studies can help to address, and refine performance measures and indicators.

Include, as appropriate, relevant personnel within and outside the agency who contribute to the accomplishment of each goal.

Support the goal leaders in assuring other organisations and programmes are contributing as expected to the goal?

Identify goals at risk of not achieving the planned level of performance and work with goal leaders to identify strategies that support performance improvement.

Encourage a meaningful dialogue around what works, what does not and the best way to move forward on the organisation’s top priorities, using a variety of appropriate analytical and evaluation methods.

Establish an environment that promotes learning and sharing openly about successes and challenges.

Agree on follow-up actions at each meeting and track timely follow-through.


While it is difficult to establish whether any particular performance system is making a difference to outcomes, a more modest goal is whether the performance data it generates are actually being used or not. Using this standard, Moynihan and Kroll (2016) found that federal government managers who took part in the development of cross-agency priority goals and quarterly reviews were more likely to report making use of performance data to take decisions, allocate resources and manage programmes. By comparison, similar tests of Bush- and Clinton-era performance reform initiatives show that they had no effect on whether managers made use of performance data. Further, the better run these quarterly reviews were – using the criteria laid out in Box 6 – the more that federal managers used data. Such quarterly reviews serve as a learning forum. Any organisation seeking to use data for learning purposes could draw on the design features outlined in Box 6 to facilitate better outcomes. Such routines can moderate the potential for negativity bias and defensiveness if they employ norms of positive reinforcement. They can also weaken
motivated reasoning by compelling people to engage in evidence-based discussion of policy implementation.

Box 6. Principles of well-run learning forums

- Meetings take place on a routine basis.
- Forums focus on important goals.
- Agency leaders are involved and seen as committed.
- Multiple levels of employees facilitate learning and problem-solving.
- Need appropriate and timely information.
- Need staff and technological capacity to analyse data.
- Quality data (reliable, accurate, valid, disaggregated to the right level, comparative) facilitate analysis.
- Follow-up on issues raised in prior meetings.
- Positive reinforcement.
- Constructive feedback.
- Reviews establish process of analysis.


New Zealand

In New Zealand, the aftermath of the 2008 financial crisis saw a need to improve public investment outcomes with fewer resources (Scott and Boyd, 2017). In 2012, the government adopted a set of inter-agency targets with ten goals of high importance to the public, which became known as the ten “Results.” Chosen by ministers, each result was given a target to be achieved by 2017, updated every six months. While some of the goals were outcomes, some were intermediate outcomes expected to generate a predictable effect on outcomes, e.g. one-stop shopping for businesses, with a target of reducing the cost of doing business with government by 25%.

Over five years there was measurable and significant improvement on all of these results, and New Zealand will adopt a new set of goals for the 2017-22 period. Scott and Boyd (2017) outline a series of lessons that contributed to the success of these cross-cutting goals. Collaboration across agencies appeared to work better when there were fewer agencies charged with managing a goal, with agencies that had more peripheral contributions given a secondary role. At the same time, they observed that the most progress was made when collaboratives leveraged existing relationships, even if a completely new approach was merited. Both insights are consistent with research on managing public sector networks (Milward and Provan, 2006).

Box 7. New Zealand’s ten Results
1. Reduce the number of people continuously receiving jobseeker support benefits for more than 12 months.
2. Increase participation in early childhood education.
3. Increase infant immunisation rates and reduce the incidence of rheumatic fever.
4. Reduce the number of assaults on children.
5. Increase the proportion of 18-year-olds with a high school diploma or equivalent qualification.
6. Increase the proportion of 25-34 year-olds with advanced trade qualifications, diplomas and degrees.
7. Reduce the rates of total crime, violent crime and youth crime.
8. Reduce the criminal reoffending rate.
9. New Zealand businesses have a one-stop online shop for all government advice and support they need to run and grow their business.
10. New Zealanders can complete their transactions with government easily in a digital environment.

In reporting to the public New Zealand anticipated some of the problems associated with numeric literacy mentioned above (Figure 1). By using graphical rates of progress, the dashboard avoids left-hand digit bias. Targets are employed, but they are not unrealistic, and rates of progress demonstrate achievement even if the target is not fully achieved. Reflecting the need to make numbers by connecting them to narratives and stories, the dashboard includes narratives that explain how the results benefit New Zealanders. In addition, agencies generated short case studies and videos, usually about how individuals engaged with these services, that illustrated these results. This content was shared with the media in an effort to generate public attention to progress.
Figure 1. How New Zealand presented progress on the Results

Lessons for managing cross-cutting goals

The US and New Zealand cases shared some lessons in common:

- **Focus on a manageable number of goals**: In the United States, just 16 cross-agency priority goals were selected for the 2015 budget. The focus in New Zealand was even tighter, with just ten goals over five years. This emphasis on a small number of goals recognises that leadership’s attention is limited and an excess of goals actually reduces the ability to make progress.

- **Participatory selection of goals**: In both cases, goals were selected based on an inter-agency dialogue in order to increase agency commitment.

- **Regular learning forums**: The United States had formal quarterly reviews to track progress. While there was no legal requirement for updates in New Zealand, inter-agency groups at multiple levels met regularly to discuss progress and strategise about next steps and a group of about 100 made up a group called “the Results Network” that prepared progress reports every six months.

- **Public commitment to targets, with regular updates on both implementation plans and metrics**: Metrics were publicly available in both countries. In the New Zealand case, the goals were written primarily for a public audience and were visible enough to capture interest from the media, making them difficult to abandon.

- **Named goal leaders**: Both New Zealand and the United States experimented with having a named leader responsible for goals. In practice, this leadership was often
shared. In the United States, White House officials shared responsibility with a senior manager of a key agency. New Zealand started with individual leaders, but evolved to a model where a group of leaders were given a collective responsibility. The gradual shift in both cases suggests that the accountability benefits of a single leader might be outweighed by the collaborative capacities that come from a team of leaders.

- **Mixture of true outcomes, intermediate outcomes and processes:** In both countries outcomes were accompanied by cross-cutting management goals that aimed to improve processes or intermediate outcomes that had a direct relationship with public outcomes. Such measures allow a sense of momentum since progress can be made in the short to medium term and they reduce the temporal problem of public investments.

- **Targets to motivate progress:** The United States focused on rates of progress rather than specific numerical targets, usually relying on narrative updates. New Zealand set more explicit targets, but avoided the bias of allowing a near-miss to be confused with failure. “In New Zealand, progress tended to be described relative to the baseline rather than the target. A huge improvement that just fell short of the target was a reason for celebration rather than punishment” (Scott and Boyd, 2017: 9).

**Conclusion: Managing both cognition and design**

Achieving some nominal level of accountability with performance data is relatively easy, if accountability is interpreted to mean simply providing more data to the public and political principals. But the ultimate goal of performance systems is to improve performance, which demands that processes of goal-based learning take place within public organisations. Learning, in turn, demands not just good data, but also an understanding of both the cognitive processes behind performance information use, as well as designing routines of performance information use. In short, it calls for both understanding what goes on inside our heads and how to change the environment around us.

In the long run, the cognitive and design aspects of performance management should be seen as complementary. Just as behavioural science argues for deliberate restructuring of processes to optimise decisions, usually referred to as choice architecture, a challenge for performance management is to minimise or avoid cognitive biases that are harmful, while exploiting cognitive tendencies in positive ways. The final section of the report makes the case that using performance data ultimately requires creating learning forums that become embedded social routines. Such routines offer a setting for structured dialogue about performance that can incorporate insights about how to minimise biases such as motivated reasoning, or amplify the positive tendencies such as our innate responsiveness to comparative data or to establish a sense of control over outcomes. Such work is more challenging than a supply-side or incentive-based approach to performance systems, since it requires investing effort in understanding where learning routines potentially exist, or could be usefully built. For example, one possibility is that ESI funds could adopt such routines via their monitoring committees, which includes representatives of the European Commission (OECD, 2016a) or in other settings. Even when such opportunities to build learning forums is identified, leaders still have to commit to their importance as social routines, and work to deliberately sidestep the types of cognitive biases that we all tend to fall into.
Notes

1. See OECD (2017a) for more on these connections.

2. Experiments are generally designed to create a high degree of confidence behind a causal claim, at least within that particular context, though they need to be replicated to feel more confident about the results. Early work on performance information use employed case studies and observational designs, where reported rates of performance information use were associated with other observed conditions, making it difficult to assert causation. Such work has been and remains important in identifying solutions the types of challenges described above. Kroll (2015) reviews this literature to identify six “important drivers of use” in performance systems. These are: 1) measurement system maturity (a well-developed performance system with available data); 2) stakeholder involvement; 3) leadership support; 4) support capacity (investment of resources into the performance system); 5) innovative organisational culture; 6) goal clarity.

3. James and Olsen (2017) provide a more detailed summary of the experimental research on how citizens use performance data, while Moynihan, Kroll and Nielsen (2017) offer more information about how elected officials and managers use performance data.

4. Observational studies have also pointed to the role of ideology in how politicians apply performance management reform and use performance data. Such a pattern became evident during the conservative George W. Bush administration in the United States, whose Program Assessment Rating Tool assessed almost all federal programmes on a five-point scale from ineffective to effective. Liberal agencies were subject to more assessments per dollar of spending, were more likely to receive lower scores, found the process more burdensome, and were more likely to be given corrective recommendations to restructure their programmes (Lavertu, Lewis and Moynihan, 2013; Moynihan, 2013).

5. In the Nielsen and Baekgaard (2015) study elected officials also provided more resources to higher performers, though this particular finding was not replicated in the George et al. (2017) study.

6. Elected officials like to call for excellence in government, and so it is difficult to find concrete examples of the type of problem-solving approach described here. One example is the “High-Risk” report series of the US legislative audit body, the Government Accountability Office, which identifies areas that will likely result in significant public failures if not addressed. Such reports draw legislative and executive branch attention to these problems.

7. My thanks to Peter Heil who made this point.
References


Moynihan, D.P. (2008), The Dynamics of Performance Management: Constructing Information and Reform, Georgetown University Press, Washington, DC.


