

Using Evidence to Make Decisions: The Experience of US Performance Management Initiatives

Donald Moynihan
University of Wisconsin-Madison

**Commissioned article prepared for the Committee on the Use of Economic Evidence to
Inform Investments in Children, Youth, and Families**

The National Academies of Sciences, Engineering and Medicine

As the ad-hoc committee examines ways in which to improve the use of data on funding decisions, it encounters a problem familiar to governments in the US at all levels. Government-wide reforms, such as the Government Performance and Results Act of 1993, the Bush-era Program Assessment Rating Tool (PART), and the current GPRA Modernization Act are prime examples of the creation of performance management systems that seek to make data more widely used in decisions. Policy-specific changes, in the areas such as welfare (the Personal Responsibility and Work Opportunity Act of 1996), education (the No Child Left Behind Act of 2002 and the Race to the Top initiative of 2009) further incentivize the use of performance measures within specific policy areas.

In my book *The Dynamics of Performance Management*, I wrote: “It is only a slight exaggeration to say that we are betting the future of governance on the use of performance information. The current era is also characterized by a willingness to adopt new structural forms of government and controls, such as networks or outsourcing, or to simply provide greater freedom to managers. New structural forms and modes of control raise difficult questions. How do we coordinate? How do we manage? How do we control? How do we exert accountability? How do we improve? How do we engage citizens? Performance information is frequently cited as the answer. We are told that performance information will allow elected officials and policymakers to set goals. It will provide the basis for accountability. It will be tied to incentives. It will allow innovations to be identified and diffused. It will enable the allocation of scarce public resources. It will allow citizens to give feedback on services. The one constant in future visions of government is the availability and smart use of performance information. If performance information does not prove to be the linchpin for the future of governance, we will have to return to the basic questions listed above, and find some alternative answers” (Moynihan, 2008, pp. 4-5).

Perhaps the clearest lesson from research in this area is that the implementation of perhaps systems are themselves complex, and subject to such implementation difficulties that they do not, cannot, neatly fill the heavy weight they bear in contemporary models of governance. There are two additional lessons to consider. One has to do with the nature of data itself, and the other is how to design learning forums that make use of data.

AMBIGUITY AND SUBJECTIVITY IN DATA

The espoused theory behind introducing more data into decisions is that it will lead to better decisions, because data is objective, standardized, indicative of actual performance, consistently understood, and prompts a consensus about how a program is performing and how it should be funded.

What I refer to as the interactive dialogue model (Moynihan, 2008) argued that performance data is not comprehensive – for any complex program or task, there are multiple ways of capturing performance, some of which will and will not be formalized. Two people could look at the same policy area, and reasonably disagree on what are appropriate measures. Performance data also has the quality of ambiguity - the same data can generate different interpretations. Data may tell us little about context, and implementation, factors that shape how we interpret whether a program is effective. Two people could look at the same data on schools, for example, and come to different conclusions both on the meaning of the data (“is the school failing or succeeding?”), and what to do next (“does it need more resources, or to be

closed down?”).

This element of ambiguity encourages subjectivity about performance, where actors associated with specific institutions, parties, or worldviews would use performance data to present the version of the world they favored. In short, performance data could not be reasonably expected to replace politics, or erase information asymmetries in the policy process. Instead, the most reasonable expectation is that data becomes part of the language in an interactive dialogue on the meaning of performance, shaped by different worldviews, interests, and power. For example, one experimental design showed that the same performance data on health services was interpreted differently depending on prior political perspectives: conservatives tended to view the same performance data more positively if told it was generated by a private rather than public provider (Baekgaard and Serritzlew, 2016).

The subjectivity of performance data is one of the reasons why Congress, despite mandating multiple forms of performance budgeting processes, has shown little interest in actually using data to make budget decisions. The potential that data will be used purposefully is more likely in homogenous settings, where individuals can agree about the basic goal of a program, and can engage in iterative discussion on the quality and relevance of data. For this reason, the different performance regimes constructed by the US government have had their greatest benefit within organizations, rather than across contentious executive and legislation forums.

Techniques such as cost-benefit analyses may seem less susceptible to subjectivity than the selection of a simple performance target. But as the importance and sophistication behind cost-benefit analysis has arisen, it has not resulted in greater agreement as to its application. As such analyses affect well-resourced groups they are apt to use those resources to contest basic premises and techniques. For example, the finance industry has developed a strategy of undermining regulation arising from the post fiscal-crisis Dodd-Frank financial reforms both by lobbying about the nature of Securities and Exchange Commission cost-benefit analyses, as well as vigorously contesting them in court (Rivlin, 2013; Scalia, 2012). While historically such analyses had been the creature of the executive branch, as they have been increasingly contested in the courts, the contested nature of data becomes clearer.

HOW TO LEARN FROM DATA

Organizational learning should be the central management purpose of performance data for complex tasks, the means by which data is actually converted into intelligent action. Learning requires a willingness to observe and correct error. This depends on frank discussions on what is working or not.

A classic error that governments have made in efforts to link data to decisions is that they establish detailed and often burdensome routines to create and disseminate performance data, but pay much less attention to creating routines to use data. Learning forums are structured routines that encourage actors to closely examine information, consider its significance, and decide how it will affect future action.

The meaning of data is not always straightforward; even the answer to such basic questions as to whether performance is good or bad may be unclear. Learning forums are important because they provide a realm where performance data is interpreted and given shared meaning. More complex questions, such as “why is performance at this level?” or “what should

we do next?” cannot be answered by looking at the data itself, but require deeper insight and other types of knowledge that can be incorporated into learning forums.

Such routines are more successful when they include ground rules to structure dialogue, employ a non-confrontational approach to avoid defensive reactions, feature collegiality and equality among participants, and include a diverse set of organizational actors responsible for producing the outcomes under review (Moynihan, 2008).

Learning forums are called different things – data-driven reviews or “stat” meetings. When the United States updated its federal performance system in 2010 it required that managers hold such meetings on a quarterly basis. While prior studies had shown that exposure to federal performance reforms had little positive effect on whether managers used performance data or not, a survey of federal employees found that those who were involved or aware of quarterly reviews about their programs were more likely to report using performance data. All learning forums are not created equal however. Among those involved in quarterly reviews, those who rated their learning forum as well-run, based on the attributes listed in table 1, were more likely to use performance data (Moynihan and Kroll, 2015).

A learning forum will be more effective if it incorporates different types of relevant information. This raises the question of what we mean by data. For example, quantitative data is more useful when it can be interpreted by individuals with experiential knowledge of process and work conditions that explain successes, failures, and the possibility of innovation (Moynihan, 2008). The latter type of information could also include some type of evaluation, ideally with treatments and controls, a cost-benefit analysis, or a cost-effectiveness analysis.

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

TABLE 1: Principles of Well-Run Learning Forum

SOURCE: From Moynihan and Kroll (2015).

While any single piece of information may be subjective and incomplete, a greater range of data offers a more comprehensive picture of the program, and a more nuanced ability to answer a greater range of questions. Performance data tells us if a certain standard was achieved. While important, knowing the level of performance should be only be one way that evidence informs the broader dialogue on effectiveness. Other types of information can answer more demanding questions. For example, experiential information can provide insights that have not been documented quantitatively, such as innovations or process changes that might have affected outcomes. Evaluations help to answer causal questions about why performance occurred and the

ultimate impact of activities on broader outcomes. Cost-benefit analyses tell us the relative cost of a certain level of performance, ideally relative to alternatives uses of tax dollars.

A perpetual struggle for the use of data in government is to include these different types of data together. In practice, mixing different type of information can be a struggle, especially as individuals are used to and have more faith in a certain type of data. A classic example of this in US government has been the failure to link program evaluations and performance metrics, which are undertaken by different staff, often with different professional training, that have come to form distinct and sometimes rival communities in government (Moynihan, 2013). Indeed, there are not many good models for how to build connections between these varied communities within government, even with the new requirement for quarterly reviews. One of the most interesting such efforts comes from outside government, with the Pew Charitable Trusts and MacArthur Foundation efforts to develop a “Results First” framework that collects evaluation and cost-benefit-analyses in certain policy areas, and has found some success when state policy analysts and lawmakers incorporate this data alongside performance information (Pew-MacArthur Results First Initiative, 2014). They have made greatest progress in examining the return on investment for prison-based programs that aim to reduce recidivism, encouraging states to shift money to less expensive but more effective techniques.

In any particular setting, the actors who make up learning forums will be different, and the range of data they deal with will vary. Any such forum will be imperfect, but creating such routines, and running them well, is a key step in fostering the use of data (Moynihan and Kroll, 2015).

REFERENCES

- Baekgaard, M., and Serritzlew. S. (2016). Interpreting performance information: motivated reasoning or unbiased comprehension. *Public Administration Review*, 76(1), 73-82.
- Moynihan, D. P. (2008). *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC: Georgetown University Press.
- Moynihan, D. P. (2013). *The New Federal Performance System: Implementing the GPRA Modernization Act*. IBM Business of Government Research Endowment. Available: <http://www.businessofgovernment.org/sites/default/files/The%20New%20Federal%20Performance%20System.pdf> [April 5, 2016].
- Moynihan, D. P., and Kroll, A. (2015). Performance management routines that work? An early assessment of the GPRA Modernization Act. *Public Administration Review*, 76(2), 314-323.
- Pew-MacArthur Results First Initiative. (2014). *Evidence-Based Policymaking: A Guide for Effective Government*. Available: <http://www.pewtrusts.org/~media/Assets/2014/11/EvidenceBasedPolicymakingAGuideforEffectiveGovernment.pdf?la=en> [April 5, 2016]
- Rivlin, G. (2013). How Wall Street defanged Dodd-Frank. *The Nation*, May 20. Available: <http://www.thenation.com/article/how-wall-street-defanged-dodd-frank/> [April 5, 2016].
- Scalia, E. (2012). Why Dodd-Frank rules keep losing in court. *Wall St. Journal*, October 3. Available: <http://www.wsj.com/articles/SB10000872396390444004704578032223012816236> [April 5, 2016]