Performance measurement as precursor to organizational evaluation capacity building

The widespread use of performance measurement and program evaluation in public administrations worldwide has been met with varying degrees of success, in terms of increased transparency and effectiveness. In many cases, the lack of impact of these functions is attributed to insufficient organizational knowledge and capacity to implement proper monitoring and evaluation systems. Both of these management tools are recognized in their own right as having the potential to contribute to ongoing decision-making and budgetary allocations within public organizations; however, they have complementary roles in terms of producing ongoing versus periodic information, and focusing on outputs and early outcomes versus longer-term program objectives. This paper attempts to bridge these two functions by proposing that performance measurement can act as a precursor to the development of organizational evaluation capacity, by providing some of the building blocks required to develop an evaluative culture within an organization. Five models of organizational evaluation capacity were analyzed to identify the extent to which performance measurement contributes to evaluation capacity building.

Service organizations around the world have been moving from rules-focused management practices to more citizen-centred public administration through various initiatives, based on the precepts of New Public Management. Hood (1991) identifies ‘explicit standards and measures of performance’ as one of the seven doctrinal features of New Public Management. This implies the definition of goals, targets, and indicators of success (preferably in quantitative terms), in the spirit of accountability for resources spent and results achieved; Results-based Management (RBM) is one such initiative. RBM essentially promotes the measurement of progress towards sought after results (or outcomes), and the modification of program design and operations based on these measures. This approach seeks to empower public managers and remove the centralized management of
resources (Aucoin, 2012). In other words, RBM is meant to improve the accountability of public institutions, and in particular, its managers, by providing accurate and timely information on program progress towards key outcomes (Aucoin, 2012; Jorjani, 1998).

Managing for results requires a comprehensive system of performance measurement1 and program evaluation. Taken together, these evaluative activities have the potential to provide a clear picture of a program’s progress towards pre-determined outcomes—‘indeed, monitoring and evaluation are seen as essential knowledge creation and management practices for any high-performing organization’ (Hunter & Nielsen, 2013, p. 8). Although the objectives of program evaluation and performance measurement are similar, their actual practices differ in nature and scope. Put simply, program evaluation periodically determines the extent to which program objectives have been achieved, and why a program or policy did or did not function as intended. Ongoing performance measurement (or program monitoring) aims to demonstrate what is going on in terms of outputs or outcomes, rather than why or how they occurred; however, it does generate questions that can be answered through periodic evaluation (Yang & Holtzer, 2006). However, in many organizations, linkages between these two functions are rare, each function planning and conducting its own work without in-depth knowledge of the other’s activities and products—even though one mechanism could often provide useful input into another (Hunter & Nielsen, 2013).

This paper, which draws upon literature pertaining to monitoring and evaluation in developed and developing nations, explores the relationship between these two functions by situating performance measurement as a precursor not only to evaluation, but to broader organizational evaluation capacity. According to Hunter and Nielsen (2013):

> Several studies have emphasized the importance of building institutional and human capacity to design and use performance information appropriately and to work within a performance-oriented organization... This often is overlooked in discussions of both evaluation and performance management, and is important at the individual, program, organizational and even societal level. (p. 15)

Indeed, organizational evaluation capacity has been of increasing interest to evaluation researchers and practitioners in recent years, mainly due to the recognition that successful evaluation production and use require ongoing capacity development throughout the organization. A brief presentation of the complementarity between monitoring and evaluation is provided next, followed by a description of organizational evaluation capacity. These elements will set the stage for our analysis of five existing ECB models and instruments, through which we aim to outline the role played by performance measurement in ECB. This analysis will yield principles that should be applied to new or revised models of organizational evaluation capacity, in order to inform further integration efforts between performance measurement and evaluation.

### Complementarity between performance measurement and evaluation

The complementary nature of performance measurement and evaluation has long been acknowledged. Hatry (2013) provides useful distinctions between these concepts; however, he emphasizes that the ‘essential purpose of both program evaluation and performance measurement is to provide information for public officials to help them improve the effectiveness and efficiency of public services’ (p. 23). Even though they may play the same overarching organizational role, each provides a type of information that the other does not provide. Nielsen and Hunter (2013) provide an interesting typology of these complementarities, based on Rist (2006) and Nielsen and Ejler (2008):

- **Sequential complementarity**, which we alluded to earlier and refers to performance measurement feeding into evaluation and vice versa
- **Information complementarity**, which refers to the fact that both functions may draw from the same data sources and reuse information for different purposes—this enables a more efficient use of organizational resources in both functions
- **Organizational complementarity**, which can be observed when the two functions are linked administratively, often in the same unit
- **Methodical complementarity**, where both functions use data collection and analysis approaches and tools
- **Hierarchical complementarity**, which refers to the use of monitoring information collected at the national or policy level towards benchmarking exercises undertaken at a more local level in an evaluation

A number of organizational benefits accrue when performance measurement and evaluation are used in complementary ways. First, where true complementarity can be found amongst all of the organizational practices and processes identified in the previous section, ‘real-time’ decision support (provided through timely performance data) enables evaluators to focus on foundational issues of contribution and results-achievement, rather than the direct measurement of outputs and early outcomes. Second, the accumulation of knowledge from both evaluation and performance data over time provides organizations with the ability to conduct meta-analyses at the policy instrument level, rather than at the individual program level. Such analyses are much more powerful
than individual evaluations and performance reports. Third, as mentioned previously, good performance measurement systems enable evaluations to be calibrated based on informational need and thus completed in a timely manner. Finally, strong performance measurement capability contributes to the development of a learning climate in the organization that is highly valuable to both performance measurement and evaluation. Decisions can be made based on evidence, the organization can demonstrate accountability and foster learning for future initiatives—all on the basis of these functions.

The complementarity of performance measurement and evaluation is therefore well established in the literature, and oftentimes, in practice. However, beyond the direct collection, analysis and use of outcome data, the actual means through which performance measurement can support the development of an evaluative, evidence-based, organizational culture are not easily identified in the current literature. The next section describes evaluation capacity building (ECB), which aims to achieve an evidence-based learning culture throughout organizations. The section also describes several models and instruments developed in recent years; the specific role of performance measurement in each of these models will be highlighted and constitutes the main focus of our analysis.

**A primer on evaluation capacity building**

Evaluation capacity building refers to the changes undertaken by organizations to integrate evaluation practice and use at all levels (Boyle, Lemaire & Rist, 1999; Cousins, Goh, Clark & Lee, 2004; Sanders, 2003; Stockdill, Baizerman & Compton, 2002). In other words, ECB ‘is an intentional process to increase individual motivation, knowledge, and skills, and to enhance a group or organization’s ability to conduct or use evaluation’ (Labin, Duffy, Meyers, Wandersman & Lesesne, 2012). Our understanding of organizational evaluation capacity (and how to build it) has evolved considerably in the past decade—we now have a better sense of how evaluation capacity develops in an organization and the ways in which such capacity can be fostered (see, for example, narrative cases by Bourgeois, Hart, Townsend & Gagné, 2011; Diaz-Puente, Yagüe & Afonzo, 2008; Haefele, Hood & Feldman, 2011; Lawrenz, Thomas, Huffman & Clarkson, 2008). We also know that organizational evaluation capacity enables organizations to not only produce high-quality evaluations, but to also use them towards program improvement and organizational decision-making (Cousins & Bourgeois, 2014). Multiple models of ECB and its organizational manifestation—evaluation capacity (EC)—have been developed based on empirical research (most notably, models and instruments developed by Labin et al., 2012; Nielsen, Lemire & Skov, 2011; Preskill & Boyle, 2008; and Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry & Balcazar, 2013). Such models and instruments aim to describe how to build evaluation capacity; how evaluation capacity manifests in organizations; and what the outcomes are for evaluation capacity. Some of these models integrate performance measurement as part of the activities, resources and systems that support organizational evaluation capacity. We were interested therefore, in exploring how each of these models and instruments address performance measurement, and how they are theorized to support the development of organizational evaluation capacity. To do so, we selected five models and instruments of EC/ECB published in recent years. We then analyzed each model to identify components related to performance measurement or ongoing data collection. Table 1 outlines the major findings of our analysis.

We were surprised to note that only one of the five models (Bourgeois & Cousins, 2013) specifically identifies performance measurement or monitoring as a component of organizational evaluation capacity. However, all five models reviewed include some mention of processes, systems, and resources allocated to data collection or broader data analysis that are akin to monitoring program outcomes. For instance, Preskill and Boyle (2008) include an ‘integrated knowledge management system’ meant to ensure that ‘the evaluation system is aligned with the organization’s other data collection systems’ (p. 456), which may include performance measurement. Further, such an integrated knowledge management system is clearly linked in the model to the development of organizational evaluation capacity ‘sustainable evaluation practice is in many ways dependent on the organization’s ability to create, capture, store and disseminate evaluation-related data and documents’ (p. 455). This supposes, for example, that performance measurement data, accessible through such a knowledge management system, could contribute significantly to evaluation by reducing the need for primary data collection as part of evaluations. Taylor-Ritzler et al. (2013) also focus on the technological resources required to conduct evaluation and their ability to support staff to compile and analyze evaluation data. Along the same lines, Nielsen, Lemire and Skov (2011) identify a technology component in their model that includes monitoring software. Although their model and instrument do not specifically include a knowledge management system, the monitoring software can support activities related to capturing, processing, and disseminating evaluative knowledge. The complementary role of performance measurement in supporting ECB is highlighted indirectly by the authors:

An ECB model for government must take into account an organization’s need to improve the capacity of ‘good governance’, through other means than solely commissioning, conducting, and using findings from program evaluation. (p. 338)
Labin and her colleagues (2012) take a broader view to organizational ECB and refer to policies, processes, and practices related to conducting and using evaluation. More specifically, their model addresses the development of an organizational culture conducive to ECB, which requires the ongoing use of data towards evidence-based decision-making. An updated version of the model published by Labin in 2014 supports this general view without further specifying the role of monitoring in ECB. This broad view is also reflected in Bourgeois and Cousins (2013), whose model addresses performance measurement in four of the framework’s dimensions: organizational resources; evaluation planning and activities; evaluation literacy; and, learning benefits. The first two dimensions refer to the organization’s capacity to conduct evaluation, while the others refer to its capacity to use evaluation. The specific components of evaluation capacity related to performance measurement in this framework are explained in more detail in the following paragraphs.

Labin and her colleagues (2012) take a broader view to organizational ECB and refer to policies, processes, and practices related to conducting and using evaluation. More specifically, their model addresses the development of an organizational culture conducive to ECB, which requires the ongoing use of data towards evidence-based decision-making. An updated version of the model published by Labin in 2014 supports this general view without further specifying the role of monitoring in ECB. This broad view is also reflected in Bourgeois and Cousins (2013), whose model addresses performance measurement in four of the framework’s dimensions: organizational resources; evaluation planning and activities; evaluation literacy; and, learning benefits. The first two dimensions refer to the organization’s capacity to conduct evaluation, while the others refer to its capacity to use evaluation. The specific components of evaluation capacity related to performance measurement in this framework are explained in more detail in the following paragraphs.

Organizational resources (the second dimension including in the framework), refers to an organization’s ability to conduct evaluation outside of the specific skills and competencies required of evaluators and organizational leaders. Two of its sub-dimensions (ongoing data collection and organizational infrastructure), refer explicitly to the existence of a performance measurement system within the organization, as well as the collection of performance data that feeds into not only evaluation, but the overall planning and reporting processes that are undertaken by the organization on a cyclical basis. Two different types of complementarity can be identified upon a closer examination of these sub-dimensions and echo some of the literature on this subject. First, there appears to be a sequential complementarity, where performance data can generate questions to be answered by evaluation studies. This is particularly relevant given the organizational infrastructure sub-dimension, where performance data feed into broader organizational planning and reporting processes, which in turn, have an impact on the types of evaluations conducted as well as the focus that they have. There is also informational complementarity at work here, where performance information is assumed to be of use and important to evaluation studies, especially at a time where evaluation calibration is becoming a greater concern related to operational efficiencies. In this sense, evaluation does not need to reinvent the primary data that have already been collected and can instead focus on the collection of additional information, as long as performance data are of sufficient quality to be considered a valid and reliable line of evidence.

### Table 1: Integration of Performance Measurement in Evaluation Capacity Building Models and Instruments

<table>
<thead>
<tr>
<th>Model/instrument</th>
<th>Performance measurement components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preskill &amp; Boyle (2008)</td>
<td>■ Multidisciplinary model that focuses on building EC and its main characteristics</td>
</tr>
<tr>
<td></td>
<td>■ Integrated knowledge management system that is complementary rather than duplicative of evaluation efforts</td>
</tr>
<tr>
<td>Labin et al. (2012)</td>
<td>■ Integrative ECB model that includes key activities and processes to build EC</td>
</tr>
<tr>
<td></td>
<td>■ Refers to policies, practices and processes required to conduct and use evaluation—performance measurement could be included here</td>
</tr>
<tr>
<td></td>
<td>■ The use of data is highlighted as one of the means through which evaluation culture can be developed in an organization</td>
</tr>
<tr>
<td>Nielsen, Lemire &amp; Skov (2011)</td>
<td>■ Conceptual model and instrument based on supply and demand for evaluation</td>
</tr>
<tr>
<td></td>
<td>■ Includes a technology component that refers mainly to monitoring software</td>
</tr>
<tr>
<td>Taylor-Ritzler et al. (2013)</td>
<td>■ Instrument based on Suarez-Balcazar et al. (2010) for use in non-profit organizations</td>
</tr>
<tr>
<td></td>
<td>■ Resources is one of the key dimensions of the instrument and includes technology required to compile information into a computerized model</td>
</tr>
<tr>
<td>Bourgeois &amp; Cousins (2013)</td>
<td>■ Conceptual framework of organizational evaluation capacity</td>
</tr>
<tr>
<td></td>
<td>■ Several dimensions specifically mention performance measurement, such as ongoing data collection; organizational infrastructure; organizational linkages; results management orientation; process use</td>
</tr>
</tbody>
</table>
Another dimension of organizational capacity to do evaluation refers to the planning and conduct of evaluations. Here, the organizational linkages sub-dimension can be considered as part of the complementarity analysis, since one of its elements refers to the co-location of the evaluation function with other key organizational areas such as performance management units. Methodical complementarity can also be found here, since both forms of knowledge production mechanisms (evaluation and performance measurement) share similar processes and tools; in addition, ongoing communications between the two corporate units, facilitated by actual physical and hierarchical proximity, can lead to a greater degree of integration and organizational efficiencies.

The overall evaluation literacy of organizational members, and especially of program managers, is considered to be an essential component of evaluation capacity. Here too, performance measurement can play an instrumental role by requiring, for instance, the development of program theory (or logic models) as well as the direct collection of performance data by the programs. Because many of these functions are usually entrusted to program managers, a greater knowledge of evaluation overall can be achieved by linking performance measurement to evaluation activities that are conducted externally. In this way, performance measurement and evaluation exhibit organizational complementarity since they are managed by different functions and individuals, yet can contribute significantly to each other’s success and influence.

Finally, the learning benefits dimension provides a behavioural description of true evaluation capacity. When organizational members start to think evaluatively across all of their activities, and when an organization values data and practices evidence-based decision-making— all types of complementarity come to light. This focus on developing an evaluation culture in the organization finds its counterpart in performance measurement:

Most assume that someone in the leadership of an organization wants to use performance measurement. The problem then becomes overcoming internal fears about ‘being measured’; ensuring integration into the organizational culture; and weaving the feedback received from performance measurement data into programmatic, personnel and budgeting decision-making. (Coplin, Merget & Bourdeaux, 2002, p. 702)

Organizations that are able to overcome such fears, whether related to performance measurement or evaluation, can then truly become learning organizations and adopt double-loop learning, which occurs when public actors test and change the basic assumptions and values that underpin their mission and key policies (Moynihan, 2005).

Interestingly, some of the elements included in the various models and instruments reviewed apply equally well to performance measurement as they do to evaluation. For example, ‘leadership’ is an element common to all five models reviewed in some manner. It is widely believed that organizational leaders must support and demand quality evaluation in order for an evaluative culture to develop within the organization. In their paper, Melkers and Willoughby (2005) also highlight the importance of leadership support for performance measurement—‘[it] necessitates a leadership commitment to recognizing the performance-related work of agency heads and department directors, in addition to simply requiring performance measurement as part of budget and management reporting’ (p. 181). Other elements included in the models analyzed, such as the technical skills of evaluators, could also easily apply to performance measurement.

Discussion

This paper sought to better understand the complementarity between performance measurement and evaluation, as seen through the EC conceptual lens. Organizations struggling to develop their evaluation capacity would do well, based on our analysis, to first look to performance measurement as a tool through which this can be partially accomplished, ‘we believe that some of the new under resourced programs lack the capacity to undertake formal evaluations and should begin by collecting information through internal program monitoring’ (Boris & Kopczynski Winkler, 2013, p. 77).

The analysis of multiple models and frameworks of evaluation capacity to identify specific elements of complementarity offers a novel way of looking at how performance measurement and evaluation intersect in an organization— at the very heart of ECB initiatives the desire to create an organizational climate that values learning and evidence-based, decision-making. This attitude must permeate all interactions that take place in the organization; it is not the exclusive domain of evaluators, but should be part and parcel of policy development and program management. Just like evaluation, the true benefit of performance measurement lies in its utilization. Also just like evaluation, questions remain as to whether or not performance data are systematically considered in organizational decision-making. As described by Coplin, Merget and Bourdeaux (2002):

Most government agencies may collect data that is [sic] or could be used for performance measurement; however, they do not have a system in place in which those data are part of decision-making processes and have not made a serious commitment to do so, whether they profess to or not. (p. 700)
Further support for the lack of use of performance measurement in organizations is provided by Hatry (2010):

Not clear is the extent to which governments have used performance information on a regular basis to help guide agency planning, budgeting, and operating management decisions. The use of performance information appears to be particularly weak, especially at the [U.S.] federal and state government levels (p. S208).

Given that all five models of evaluation capacity integrated at least one element related to program monitoring or ongoing data collection, we would surmise that performance measurement can act as a precursor to evaluation capacity. The explicit inclusion of performance measurement in models of organizational evaluation capacity is, in our opinion, essential. Such models should consider the following elements (among others): developing program theory/logic models; collecting quality data related to outputs and early outcomes; integrating measurement of efficiency and sustainability; and, keeping program staff and managers engaged by involving them in monitoring efforts.

Our analysis also leads us to believe that it is by better understanding the role of performance measurement in developing an organization’s evaluation capacity, both performance measurement and evaluation can be enhanced and improved. However, even though performance measurement could contribute in many ways to building evaluation capacity, complementarity does have its limits, and depends greatly upon the extent to which performance measurement has been established in the organization:

The challenges are for a large part operational. On the one hand the challenge to select appropriate techniques to gather data persist. On the other hand, once responsibility, purposes, lines of command, and so on, are determined, M&E efforts become inscribed in wider institutional and political contexts and contests. (Lahey & Nielsen, 2013, p. 55)

Therefore, it seems just as important and critical for organizations to build their performance measurement capacity as it is for them to build their evaluation capacity, and indeed, there are likely many parallels between the two types of capacity. For instance, Moynihan (2005) discusses the implementation of performance measurement and its link to organizational learning, much in the same way as others have presented ECB. In the past, ‘reforms have occurred without the additional, necessary changes to organizational structure and environment’ (p. 203). Moynihan identifies the learning problems encountered by performance measurement implementers. As in our own framework of EC, he points out that the successful implementation of performance measurement requires a focus on double-loop learning, exemplified through a ‘broad understanding of policy choices effectiveness’, which we identify as process use (Amo & Cousins, 2007). Also noted is the importance of cultural approaches to learning, which are, in Moynihan’s view, essential to the successful implementation of performance measurement systems. Mayne (2007) further elaborates:

An evidence-based outcome focus can require significant and often fundamental changes in how an organization is managed…it requires behavioural changes…and it usually requires that elusive ‘cultural change’, whereby performance information becomes valued as essential to good management. (Mayne, 2007, p. 89)

Just as performance measurement has been shown to contribute to ECB, the opposite is also true. According to de Lancer Julnes (2013), evaluation can contribute to monitoring capacity in a number of ways including:

- increasing citizen engagement in performance measurement by using participatory approaches to community-based measurement
- understanding and developing program theory, which includes elaborating models to identify key assumptions and risks
- understanding systems theory to avoid oversimplifying complex causal relationships

In addition to these, evaluators can specifically contribute to and support performance measurement activities by ‘giving seminars on data analysis and reporting’, as well as ‘helping program managers understand to what extent their outcomes can be attributed to the program’ (Johnsen, 2013, p. 99).

**Conclusion and next steps: building organizational evaluation and performance measurement capacity**

Performance measurement has the potential, as an organizational practice, to positively contribute to the development of an evaluative culture focused on evidence-based decision-making. The evaluation capacity models and instruments reviewed in our analysis reveal for the most part, at least some consideration of the role of performance measurement in ECB. Although the evaluation literature focuses almost exclusively on ECB without clearly distinguishing between evaluation and performance measurement, the results of our analysis suggest that an organizational strategy focusing on overall data valuing and evidence-based decision-making would support the development of both functions. Based on Mayne (2007), Hunter and Nielsen (2013) suggest ways in which both evaluation and performance measurement capacity can be improved including: strong leadership that values evidence-based information and provides the right incentives; realistic expectations about the value of...
performance measurement and evaluation; supporting and encouraging innovation and organizational learning; fostering data quality and timeliness; and, transparency in reporting to keep staff, management, policymakers, and citizens engaged. Evaluation has already made some significant strides in this regard; some of the lessons learned along the way could inform more specific performance measurement, capacity building initiatives, or provide the necessary context for joint capacity building throughout the organization. For instance, it may be interesting to move beyond individual programs and shift our communal focus as evaluators and performance measurement experts to broader policy instruments and program archetypes. This would enable greater reach across organizations, thus increasing our common evaluation and performance measurement capacity as we communicate successes and share our lessons learned.

Endnote

1. For the purposes of this paper, performance management refers to the ongoing collection and analysis of output and outcome data, meant to support an organization in decision-making and goal attainment. We use the term ‘monitoring’ interchangeably.

Acknowledgments

Our sincere gratitude to the anonymous reviewer whose considerate comments and suggestions strengthened this paper immeasurably.

References


