STELIC

Measuring the Quality of Management in Federal Agencies

Measuring the Quality of Management in Federal Agencies

Table of Contents

One interviewee commented, 'Often the reason that programs fail is because of bad management rather than because of bad policy. It is hard to get people to pay attention to management.' The creation of a measure or set of measures would help bring focus not only to the management function and to those who participate in that function but to the question of how to achieve management excellence.

Introduction

Is the Department of State well managed? How about the Departments of Agriculture, Interior, or Defense? Importantly, where would one look for data to answer such a question.

- The annual Best Places to Work rankings compiled by the Partnership for Public Service?
- Measures of the job-related attitudes of employees from the Federal Employee Viewpoint Survey?
- The Government Accountability Office's list of "high risk" programs?

Such data provide valuable insights into the quality of an agency's management but neither individually nor collectively capture the full scope and complexity of the management function.

One could review each agency's annual performance report. Did the agency meet its objectives? How did the agency perform against its targets? Are key performance indicators trending in the right direction? Here too, however, the picture is incomplete. An agency may fail to meet its performance targets not because it is poorly managed but as a consequence of factors beyond its control, such as budget cuts. High quality management is a precursor of good performance but does not equate to good performance.

The propositions explored in this special report are premised on two widely-held assumptions:

- First, it is possible to develop a valid measure or set of measures of agency management quality.
- Second, there is significant value to be gained from such an exercise.

The authors—and many good government advocates—believe that agencies could benefit from a gauge of how well they are doing in comparison with others as well as from information on their strengths and weaknesses. Agency managers would gain knowledge and insight from a conversation around the development and ongoing use of a measurement protocol. The American people would benefit from a government that includes high-performing organizations permeated by an ethic of management excellence that a system of measures would promote.

Why Measure the Quality of Management in Agencies?

The challenge of developing a valid measure, or set of measures, of the quality of management in any large organization is daunting—in the public as well as in the private sector. The management function in all large organizations encompasses a broad array of objective and subjective activities. The existing inventory of management-related metrics is vast. Should the focus be on metrics and the development of a set of "meta-measures" that capture the essence of the management function? Or should it be on the identification of some set of practices that distinguish well-managed organizations from less well-managed organizations?

Notwithstanding the magnitude of the conceptual, methodological, and organizational challenges, the prospective benefits of a program to measure management quality would be great. One such benefit is the focus that would be placed on the management function and on the centrality of that function to good performance. One interviewee commented, "Often the reason that programs fail is because of bad management rather than because of bad policy. It is hard to get people to pay attention to management." The creation of a measure or set of measures would help bring focus not only to the management function and to those who participate in that function but to the question of how to achieve management excellence.

Another benefit would be the learning dynamic that such an exercise would provoke. Agencies that score well would become exemplars from which others can learn. Dialogue that would be generated around the question "What are the practices and habits of thought that characterize well-managed organizations?" would itself serve learning purposes. Such dialogue could occur not only within agencies, but between agencies and central staff units and across the federal management community generally.

Experts who participated in a November 2019 workshop to help develop this special report pointed out that federal managers want to do a good job and that many would take advantage of an opportunity to learn more about good management and about the practices in which well-regarded managers and organizations engage. An exercise such as that envisioned here would strengthen and promote the ethic of professionalism that these managers embody. In fact, only if individual managers from throughout the executive branch perceive such a program as promoting professional values would they want to be involved. Only if they were involved could such a program be sustained, and only if it is sustained could the program have a long-term impact.

The ultimate payoff of an initiative to measure management quality would come in the form of improved program outcomes for citizens and businesses. An initiative to measure management quality is by definition holistic in nature, relating as it must to all aspects of how an agency functions and to all levels of an organization's hierarchy. A broad and inclusive perspective of management quality would in turn be highly conducive to a range of positive outcomes—service quality, financial soundness, employee engagement and development, and above all, program effectiveness and mission accomplishment.

Methodology

This special report was prepared subsequent to the workshop on "Measuring the Management Quality of Federal Agencies," hosted by the Stelic Center for The Business of Government in Washington, D.C., on November 20, 2019, and co-sponsored by the Senior Executives Association and the University of Illinois—Chicago.

The purpose of the workshop was to obtain the ideas and insights of the participants on the design of an initiative to measure the quality of management of federal agencies. At the workshop, the authors of this special report presented research on several initiatives that have a similar purpose in the private as well as the public sector.

The authors also summarized what they had learned on this topic from a series of interviews with individuals with expertise in the management of federal agencies. This report serves as a synthesis of what has been learned from the research, the interviews, and the workshop in the form of a summary of key issues and ways forward.

The Time Is Right

The administrative routines and the data infrastructure created to implement the Government Performance and Results Act of 1993 (GPRA) and the GPRA Modernization Act of 2010 have contributed to a growing level of sophistication in the development and use of management-related metrics. These laws and other related data collection efforts have resulted in the availability of rich data sets on agency management metrics. For example:

- Measures tracked under the "Developing a Workforce for the 21st Century" cross-agency priority (CAP) goal include several based on the Federal Employee Viewpoint Survey, as well as one on new hire retention rates and one on delegated examination certification and training.
- In 2019, 25 "high-impact service provider" agencies began reporting data on customer experience, including measures of customer satisfaction, issue resolution, and whether the customer felt fairly treated.
- Under the, "Sharing Quality Services" CAP goal, data is reported on manager satisfaction with support services in the areas of human capital, financial management, contracting, and information technology.

However, a key challenge in developing an initiative to measure agency management quality is to devise a framework within which these and other metrics can be meaningfully synthesized and interpreted.

What Has Been Learned About Measuring Management Quality?

In addressing such a challenge, advantage can be taken of the learning that has already occurred around the measurement of management quality in the private sector. For example, researchers from Harvard, Stanford, and MIT developed the World Management Survey (WMS) which incorporates a list of 18 "core managerial practices," associated with well-run organizations. The WMS team interviews mid-level managers to assess the extent to which such practices have been institutionalized in each of the organizations included. In a 2017 article in the *Harvard Business Review,* the principals involved in this initiative presented data showing a high correlation between the adoption of these practices and firm performance (Sadun, Bloom, & Van Reenen, 2017).

Similarly, the Baldrige Excellence Framework—a public-private partnership hosted by the U.S. Department of Commerce—is designed to help organizations in the private, nonprofit, and public sectors identify management strengths and weaknesses, to learn successful management strategies and practices, and to thereby achieve performance excellence. The Framework includes over 80 separate items in the areas of:

- Leadership
- Strategy
- Customers
- Measurement, analysis and knowledge management
- Workforce
- Operations
- Results

Other efforts to measure management quality in the federal government have included the Federal Performance Project (FPP) and the Program Assessment Rating Tool (PART).

The FPP was a nongovernmental collaborative effort between the Pew Charitable Trusts, George Washington University, and *Government Executive* magazine whereby grades were assigned to federal agencies in the areas of:

- Financial management
- Human resources management
- Information technology management
- Capital management
- Managing for results

The FPP initiative operated between 1998 and 2001.

PART was a governmental project that was part of a two-part rating system included in the 2001-2009 President's Management Agenda (PMA). In the first tier, each department's key management functions were rated as red, yellow, or green on each of the five elements of the PMA:

- Strategic management of human capital
- Competitive sourcing
- Improved financial performance
- Expanded electronic government
- Performance improvement

Within the "performance improvement" component, there was a separate rating system for individual government programs called PART. Pursuant to PART, examiners at the Office of Management and Budget (OMB) assessed over 1,000 individual federal programs on the basis of program purpose and design, strategic planning, program management, and program results (Congressional Research Service, 2009).

Two more recent federal statutory management reform initiatives to improve management quality are not focused on measurement per se but would complement a measurement program:

- The GPRA Modernization Act (GPRAMA) of 2010 requires agencies to conduct regular, data-driven performance reviews as a means of improving program outcomes and of promoting the use of data for decision-making purposes. Such reviews constitute the type of management "best practice" that a measurement program would help identify and promote.
- The Program Management Improvement and Accountability Act (PMIAA) of 2016, in turn, addresses the need to increase management capacity at the program level. It is at that same level that management quality ultimately becomes manifest in the form of high-quality program outcomes for citizens and businesses.

Examples of initiatives that use measurement as a means of promoting management quality can be found in other countries as well:

- In 2005, the United Kingdom launched the Capability Assessment Review (CAR) program designed to "assess systematically the organizational capabilities of individual departments and to publish results that can be compared across departments" (National Audit Office, 2009, p. 4). Pursuant to that program, separate sets of external reviewers rated each of 17 departments in the areas of leadership, delivery, and strategy on the basis of a protocol developed by the Prime Minister's Delivery Unit. Although the CAR program was phased out subsequent to a change in government in 2010, Australia and New Zealand implemented their own capability assessment programs based on the CAR framework (NAPA, 2018).
- In Canada, pursuant to the Management Accountability Framework (MAF), agencies submit data in response to an interrogatory from the Treasury Board Secretariat (TBS). The MAF is used to gauge agency compliance with TBS policy directives, as a means to assess the performance of deputy department heads and to "develop a comprehensive and integrated perspective on management issues and challenges" (Treasury Board Secretariat, 2017).[1]

1. Please review Appendix B for more details about the instruments mentioned in this section.

Six Key Building Blocks

Following are six key building blocks that could serve as a foundation for designing and implementing a governmentwide initiative to measure the management quality of federal agencies.

Building Block #1—Have the Design of a Measurement Protocol Serve as a Learning Opportunity

Initiatives such as those cited above confirm that it is possible to devise a valid measure of management quality. However, there are a large number of design issues to be addressed including the following: Who would sponsor such an initiative? How would one go about identifying a set of metrics that capture management excellence? Should a qualitative element be included and what form would that element take? Would the measurement protocol have to be customized to an agency's size or the nature of an agency's work? What key practices distinguish well-managed agencies from less well-managed agencies?

Key to the present concept are: 1) The process of addressing such questions could itself serve important learning purposes, and 2) the initiative should therefore be structured so as to promote communication and dialogue that can serve learning purposes. Modern communication technologies could be employed as a means of facilitating broad participation in the design of a measurement protocol.

Next Steps:

- Government leaders and non-profit partners should create a temporary working group to assess interest in a program to measure the management quality of federal agencies.
- Based on existing instruments and approaches, develop a draft measurement protocol.

- Pilot the protocol with a few select agencies.
- Seek guidance and advice from internal stakeholders such as OMB, the Government Accountability Office (GAO), the Office of Personnel Management (OPM), and the General Services Administration (GSA), as well as from external stakeholders such as the Senior Executives Association (SEA), the National Academy of Public Administration (NAPA), the Partnership for Public Service (PPS), the Federal Managers Association, the Professional Managers Association, and the federal employee unions.

Building Block #2—Balance "Bottom-Up" with "Top-Down" in the Design of a Measurement Initiative

For a measurement initiative to have an impact, it will need to be sustainable. To be sustainable, ownership will need to be shared including by managers and executives from throughout the executive branch. Such considerations imply a "bottom-up" approach with participation by a wide range of stakeholders.

Notwithstanding reliance on a bottom-up philosophy, there would also be a need for central staff involvement to legitimize and promote a quality measurement initiative. A challenge will be to find the right balance with OMB, OPM and GSA participation but with a focus on learning rather than compliance. Central staff domination could inhibit the growth of a sense of ownership on the part of agency management. A comment made with regard to New Zealand's centrally-administered Performance Improvement Framework (PIF) is revealing in this regard;

> Is the PIF about holding agencies publicly accountable or is it about organisational improvement. . . . although they are told otherwise, many, especially below CEO level, believe that the primary purpose is organisational accountability. They feel marked, graded, and compared, rather than motivated to lift their strategic goals and be more ambitious in what they are trying to change and achieve at the outcome level" (School of Government, 2017, p. 30).

Considerations of sustainability also mitigate in favor of an agency- and manager-driven approach. Centrally-sponsored management improvement programs are susceptible to termination upon a change in administration. In their review of the CAR program, Panchamia and Thomas (2014, p. 8) explain that in the U.K., "The Conservative opposition saw capability reviews as part of Labour's top-down performance management approach, which it was keen to challenge." In the U.S., the PART program experienced a similar fate upon the transition from the Bush administration to the Obama administration. For a measurement program to have an impact and for the learning dynamic to play out, a long-term orientation is key.

Where would OMB fit? On the one hand, OMB's traditional role has been more top-down than bottom-up. On the other hand, OMB is uniquely positioned to legitimize a measurement initiative and to induce agency participation. One means of achieving that purpose would be to integrate a quality measurement program with other management improvement initiatives. Could, for example, a measurement initiative be integrated with the strategic review process set forth in OMB's guidance Circular A-11?[2]

Participants in the November 2019 workshop endorsed the idea of a "hybrid" model with the involvement of an outside entity subject to inside direction. Examples include a federally-funded research and development center (FFRDC) and an entity created pursuant to the Federal Advisory Committee Act (FACA). An advantage would be that while OMB would remain a sponsor and participate in the design, execution could take place outside the federal establishment thereby enhancing prospects for sustainability.

Next Steps:

- Engage OMB in a discussion of how a measurement initiative could be incorporated in the A-11 framework.
- Design a governance structure to identify the roles of central staff units, agencies, executives, managers, and manager groups.
- Assess the viability of assigning program execution to an outside entity.

Building Block #3—Engage the Federal Management Community

A workshop participant endorsed an "organic" approach to measuring management quality. The word "organic" implies an approach whereby the program would start small and grow over time as new knowledge is acquired, as different groups and parties have a chance to weigh in, and as the management function itself evolves in response to technological, demographic, and environmental developments. It implies an evolving rather than a fixed format.

Open innovation techniques such as crowdsourcing would be congruent with such a model. Such techniques operate on the premise that the quality of the outcome correlates with the breadth of participation. Thus, for example, managers from throughout the executive branch could be invited to participate in the design of a protocol to measure the management quality of federal agencies and hence to shape the product. Such an approach would contribute to a high level of "buy-in" and hence to program sustainability.

A "community of practice" around management quality and management excellence could emerge. Such a community could include rank-and-file managers, supervisors, and executives. Among the topics on which such a community could focus would be the continued refinement of the measurement protocol and what workshop participants described as a "playbook" of managerial best practices compiled in the course of the measurement process. Managers in agencies subject to an assessment or review would have particular incentives to participate and to contribute to the development of a protocol that would accurately reflect their agency's strengths and weaknesses.

2. Circular A-11 is issued annually by OMB as a guidance for federal agencies in several policy areas. Specifically, Part 6 of Circular A-11 focuses on improving the management and organizational performance of federal agencies (Performance.gov, 2019).

Too often, programs to improve management practices have had limited penetration within agencies. The intent of a measurement program would be to engage managers at all levels and in all types of functions. In fact, it is at the program level and with those responsible for program delivery that management quality, or lack thereof, becomes manifest. The contributions of these individuals to the design and implementation of a measurement program would be of particular importance. The contribution of staff units such as human resource management and procurement to overall management quality should be gauged based on how well those functions serve the needs of program personnel. The importance of effective integration of support services and program management was attested to by one interviewee in the context of the IRS's efforts to combat identity fraud.

> I want our HR (human resources) people to know that one of our top priorities is to rid the tax system of identity theft. If, when my hiring people are preparing position descriptions, they are completely divorced of what the skills and capabilities needed to combat identity theft, then my HR function is disconnected from my mission.
>
> —*IRS Interviewee*

Next Steps:

- Encourage broad participation in the design of a protocol to measure agency management quality.
- Explore the use of open innovation techniques such as crowdsourcing for design purposes.
- Encourage managers from throughout the executive branch to share knowledge, insights, and best practices as a community of practice.

Building Block #4—Use the Measurement Initiative to Rethink Management

Several of the measurement initiatives cited above conceptualize management in a functional framework—HR, IT, finance, procurement, capital assets, etc. While these functions represent core management activities, whether an organization is well run or is assessed ultimately based not on how each of multiple individual functions operates but on the degree of integration that is achieved internally across functional units and between support functions and program management and externally as well. Much progress has been made, for example, in leveraging resources and expertise across agency boundaries through use of the shared services model. OMB has encouraged an executive branch-level focus through the identification of cross-agency priority goals. The cross-agency councils comprised of different mission support leaders (e.g., chief information officers, chief financial officers) have promoted cross-agency communication and learning. Good management practice requires that such partnering and networking activities become an integral part of how each agency does business. A measurement protocol therefore would need to address this collaborative element.

Central to such an element would be the engagement of the public in the execution of agency responsibilities. The Government Accountability Office (GAO, 2017) has identified five "practices to engage citizens and effectively implement federal initiatives," including crowdsourcing and citizen science, idea generation, open data collaboration, open dialogue, and prize competition or challenge. Regulations.gov serves as a model of how agencies could constructively engage the public in pursuit of their missions. Other open innovation initiatives cited by GAO include openFDA and the My Air, My Health Challenge sponsored by the National Institute of Environmental Health Services, the Environmental Protection Agency, and the Department of Health and Human Services. Such initiatives are illustrative of how the knowledge, experience, and expertise of the general public can be leveraged for governance purposes.

Next Steps:

Assess agency management on the basis of whether and to what extent:

- Agency support services meet the needs of program managers.
- The agency partners with external entities in the pursuit of its mission.
- The agency engages the general public in pursuit of its mission.

Building Block #5—Evaluate the Advantages and Disadvantages of Assigning Agencies Grades

Apparent from the experiences of other countries is that whether or not to assign grades is a critical design question. Canada started but stopped assigning grades in part as a consequence of agency resistance. A danger is that publicly-reported grades cause agencies to adopt a defensive posture at odds with an attitude of openness needed to serve learning purposes.

Grading further tends to induce gaming behaviors. In his review of the PART program, Gilmour (2006) comments, "Officials from several of the programs that raised their ratings explained their success with almost identical language. 'We learned how to take the test,' they said" (p. 17). Gilmour adds, "None of the officials interviewed for this project claimed that they had introduced significant management improvements or changed program design in order to raise their rating" (p. 18).

A counterargument is that absent consequences for a finding of poor management quality, agencies are less likely to divert the time, energy, and resources necessary to achieving management excellence. An excerpt from an evaluation that was done of the U.K.'s CAR program is relevant in this regard:

> The reviews came to be seen as robust primarily because they applied the model consistently across all departments and allowed for comparison between them. This focused the attention of permanent secretaries on the relative strengths and weaknesses within their departments, and injected a degree of competition, which acted as a constant pressure for improvement: 'Without the scores, they wouldn't have listened'" (Panchamia & Thomas, 2014, p. 5).

A “middle ground” would be to assign grades but to seek ways of mitigating any negative impacts. In the U.K., Australia, and New Zealand, grades (such as Strong, Well Placed, Development Area, Urgent Development Area, and Serious Concerns) are assigned in each of three areas of focus—leadership, delivery, and strategy—but there are no overall grades that would enable outsiders to make simplistic comparisons between agencies.

A second means of mitigating the adverse consequences of grades would be to employ a maturity model with an emphasis on the developmental purposes of the exercise. Thus, for example, each management dimension (e.g., human resource management) could be assigned one of four maturity ratings based on a specific set of criteria associated with each stage of maturity. OPM uses a maturity model (namely, Reactive, Emerging, Advanced, Optimized) in conjunction with its HRStat program (OPM, 2018) as does OMB with regard to agency strategic reviews (GAO, 2015).

A third means of mitigating the adverse consequences of grading would be to focus on the future. The question would become, “How well is the agency positioned to deliver now and in the future?” (New Zealand State Services Commission, 2015, p. 3), rather than, “How have you done in the past?” In its evaluation of New Zealand’s Performance Improvement Framework (PIF), a team from Victoria University commented, “The PIF’s focus on the four-year excellence horizon is widely accepted and is seen to have pushed organizations toward more strategic and longer-term thinking” (School of Government, 2017, p. 28).

One interviewee recommended that grades be phased in over time, that the program sponsor work for a year or two with the organization and give the agency a chance to “do some learning. You would tell the organization that by year three it would be made public. It gives the agency time to ramp up. Agencies will be hesitant if it is going to be splashed everywhere.”

Next Steps:

- Decide whether or not to issue grades in conjunction with the measurement protocol.
- If grades are to be included, determine what framework will be employed (e.g., maturity model).
- Define the outcome of the measurement process and whether that outcome is to be made public.

Building Block #6—Determine the Degree of Reliance to Be Placed on Metrics

The “Time Is Right” section above lists several of the management-related metrics that are being collected across the government and that could be incorporated into a protocol to measure agency management quality. A question that arises in this context is to what extent management quality can be assessed exclusively or primarily on the basis of metrics—or, alternatively, would such a protocol need to incorporate a qualitative element such as interviews, focus groups, or surveys.

The metrics cited above including those that derive from the Federal Employee Viewpoint Survey (FEVS), the data collected by the GSA on manager satisfaction with support services, and the data on customer experience compiled by OMB represent but a small subset of the universe of available metrics. For example, in its HRStat Guidance, OPM lists 15 metrics that derive from the FEVS and another 21 that derive from other sources. Data on agency information technology projects is available on the dashboard.gov website. Data on improper payment amounts and rates is posted on the paymentaccuracy.gov website. Amidst this abundance, which are those "vital few" metrics that most closely correlate with overall management quality?

One approach to making this determination is to assess which of the different management-related metrics correlate most closely with high quality program outputs and outcomes. As one example, the Partnership for Public Service and the Boston Consulting Group (2019) found a positive correlation between patient satisfaction and employee engagement at the Department of Veterans Affairs. Another approach would be to categorize metrics by tier where those assigned to the secondary and tertiary tiers are determined to be driven in part by those assigned to the primary tier. An early step would be to inventory the existing requirements, mechanisms, and databases currently being used for measurement purposes by agencies.

As necessary as analyses of this type would be, a perception among workshop participants and interviewees is that, in assessing management quality, it would need to be supplemented with a form of qualitative data. One interviewee commented, "The data doesn't tell you the answers, it tells you the questions." Along similar lines, an interviewee commented that it is not whether an agency has a strategic plan or a staffing plan but, "Do you manage with it on a day-to-day basis?" Such a determination is difficult to make absent direct engagement with management personnel.

A second set of questions relates to the form that a qualitative element would take. Options include interviews with senior managers, focus groups, and/or surveys of mid- and upper-level managers. In the U.K., agency reviews were conducted by separate teams each consisting of two executives from other government agencies and three from outside organizations. Each review team assigned ratings in the three areas listed above on the basis of its review of materials submitted by the agency and of a series of interviews with senior leaders. Evaluations of the CAR cite a generally positive disposition toward the reviews on the part of agencies: "The review team members' interaction with the permanent secretary and the board is crucial in ensuring the review's conclusions are both valid and especially accepted by the top management team" (Sunningdale Institute, 2007, p. 12).

In its evaluation of New Zealand's Performance Improvement Framework (PIF), a team from Victoria University commented, "Much of the creditability, longevity, and sustainability of PIF starts with having perceptive lead reviewers who are considered to be highly respected by relevant agencies" (School of Government, 2017, p. iv), and further that, "A number of case study respondents specifically mentioned the value that was to be gained from speaking with a lead reviewer. Respondents indicated they thought that knowledge and systemwide understanding the reviewers brought to their PIF was invaluable" (School of Government, 2017, p. iv).

It is also the case, however, that the use of reviewers and review teams introduces complications and costs. There would be a need for example to identify a means of standardizing ratings across review teams in order to mitigate concerns about subjectivity. There would be a cost associated with recruiting external reviewers.

Next Steps:

- Inventory the metrics available for the purpose of measuring management quality.
- Decide whether to include a qualitative element and if so, the form that element will take.
- If the protocol is to include reviewers, identify potential pools of candidates.

Conclusion

The development and implementation of a protocol to measure the management quality of federal agencies is a potentially powerful means of promoting good management with a variety of potentially valuable outcomes. The first and most obvious is simply to be able to identify which agencies are well-managed and which are less well-managed. What are the factors that account for those outcomes? What are each agency's managerial strengths and weaknesses?

A second key outcome would be the learning dynamic that would result with the identification of specific practices in which well-managed agencies engage and which others could potentially replicate. The conversation that would take place around the creation of a measurement protocol would itself serve as an important source of learning: What are the core practices that distinguish well-managed organizations from less well-managed organizations? What are the core few metrics that best correlate with high-quality management?

Importantly, if the initiative is to last, the ownership needs to be distributed rather than centralized. Distributed ownership, in turn, implies that managers and executives from throughout the executive branch have a say in how the protocol is designed and implemented. At the same time, the prospect for such an initiative coming to fruition would be enhanced to the extent it was integrated with other management improvement initiatives including those relating to GPRA, GPRAMA, the FEVS, and the President's Management Agenda. The learning purposes served would synchronize well with the mandate included in the Foundations for Evidence-based Policymaking Act for agencies to develop learning agendas.

An initiative premised on broad involvement and bottom-up design would represent something of departure from business as usual in the executive branch. At the same time, there would be great potential in providing a new outlet for the application of the knowledge, expertise, and commitment to excellence with which the management cadres in agencies throughout the executive branch are replete.

Appendix 1

Interviews Conducted

From June to September 2019, 37 interviews were conducted with the aim to identify and understand the metrics used by agencies that serve as indicators of management quality. Additionally, the interviews had the purpose of learning each interviewee's perceptions about the viability of a program to measure management quality and about how such a program should be designed. Some of the questions asked included: Can a protocol resulting in a valid measure of the management quality of federal agencies be devised? Should such an initiative be hosted from within or outside the government? If hosted from outside the government, how would agencies be induced to participate? How could such an effort complement or incorporate existing oversight activities and thereby mitigate the burden on agencies?

The interviews were conducted with current and former government managers and executives as well as with individuals from the private sector and from other countries. The distribution of the thirty-seven interviewees according to sector is as follows:

- Federal—20
- Former Federal—9
- Nonfederal—3
- Non-U.S.—5

The interviewees from the federal government represented the following agencies:

- Office of Management and Budget
- General Services Administration
- Government Accountability Office
- Consumer Protection Financial Bureau
- Department of Defense
- U.S. Agency for International Development
- Department of Homeland Security
- Department of the Interior
- National Institutes of Health
- National Aeronautics and Space Administration
- Department of Veterans Affairs
- Office of Personnel Management
- Department of the Treasury
- Department of Housing and Urban Development
- Small Business Administration

Appendix 2

Existing Instruments Designed to Measure Management Quality

World Management Survey

The World Management Survey (WMS) project started in 2002 as a means of measuring managerial quality. The WMS instrument is based on an in-depth survey methodology designed by the McKinsey Company in early 2000's. The instrument was used initially to analyze differences across firms and countries in the manufacturing sector but has subsequently been expanded to include hospitals, schools, and retail outlets. The basic WMS methodology contains 18 questions within four areas (Bloom, Van Reenen, 2007):

- Operations management
- Performance monitoring
- Target setting
- Talent management

On the basis of the information gathered, the sponsors rate each organization on each management practice, using a one to five scale in which higher scores indicate greater adoption of each management practice. The ratings are averaged to produce an overall management score for each company (Sadun et al., 2017).

The WMS website allows users to benchmark their own organizations through a free, web-based tool. After answering a set of questions, the user's answers are compared with others contained in the WMS database to thereby enable benchmarking. Currently, the methodology has been implemented in 35 countries and 20,215 interviews have been conducted (World Management Survey, n.d.).

One of the main conclusions of the WMS is that measures of management quality are positively associated with better firm performance (Bloom and Van Reenen, 2007). Variables that have been found to explain differences in management practices across companies and countries include product market competition, labor market regulations, ownership, and meritocratic selection of the CEO, multinationals and exports, human capital, and information (Bloom and Van Reenen, 2010).

Baldrige Performance Excellence Program

The Baldrige Performance Excellence Program operates as a partnership between the U.S. National Institute for Standards and Technology and the Baldrige Foundation. The program is designed for organizations in both the public and private sectors to assess themselves in the areas of:

- Leadership
- Strategy
- Customers
- Workforce
- Operations
- Results
- Measurement, analysis and knowledge management

The program provides a scoring guide for this purpose. An organization can, if it wishes, apply for the Malcolm Baldrige National Quality Award. Award program participants are subject to a review by a team of outside reviewers. Up to 18 awards are given annually across the following six categories—manufacturing, service, small business, education, health care, and nonprofit.

Drucker Institute's Management Top 250 Ranking

The Drucker Institute compiles the Wall Street Journal's annual, "Management Top 250," which includes the 250 companies judged to be the best managed. The ranking program is based on five core principles of corporate performance:

- Customer satisfaction
- Employee engagement and development
- Innovation
- Social responsibility
- Financial strength

Effectiveness in these areas is assessed on the basis of 37 separate indicators. The Top 250 ranking was first published in 2017. This methodology is intended to measure corporate effectiveness and performance, rather than management quality per se (The Drucker Institute, n.d.).

U.K.'s Capability Assessment Review Program

The Capability Assessment Review Program (CAR) started in the U.K. in 2005 and was subsequently emulated by Australia and New Zealand. Currently, the CAR is only in place in New Zealand. The main purpose of this program was to hold departmental leaders accountable for improving their capability to deliver and meet future challenges. This instrument addressed three management areas: delivery, strategy, and leadership.

Administered by the Prime Minister's Delivery Unit, the CAR included review teams with executives from the governmental sector and external members from the private and voluntary sectors. It had a rating system for each of the three management areas with five different rating categories: Strong, Well Placed, Development Area, Urgent Development Area, and Serious Concerns (Panchamia & Thomas, 2014, National Audit Office, 2009).

Like the WMS instrument, the CAR included qualitative data obtained through interviews with senior managers. The reports were publicly available.

This instrument focused on 10 elements of capability organized around the three main management areas: leadership, strategy, and delivery. The 10 elements included:

- Set direction
- Ignite passion
- Pace and drive
- Take responsibility for leading delivery and change
- Build capability
- Plan, resource, and prioritize
- Develop clear roles, responsibilities and delivery models
- Manage performance
- Build common purpose
- Base choices on evidence
- Focus on outcomes

The reviews were conducted at the departmental and agency levels (NAPA, 2018).

Canada's Management Accountability Framework

Canada's Management Accountability Framework (MAF) is a program administered by the Treasury Board Secretariat (TBS), whereby federal departments are subjected to an annual assessment of their performance in seven areas of management, including:

- Financial management
- People management
- Information management and information technology management
- Results management
- Management of acquired services and assets
- Security management
- Service management

The MAF started in 2003 and is intended to serve several purposes including to promote management excellence, to promote cross-department learning, and to provide a basis to assess compliance with TBS directives.

Each of the 36 large departments is required to provide data to the TBS on a set of indicators in each of the areas of management. A team at the TSB reviews each department's submission and produces a report for the Secretary of the Treasury Board who, in turn, sends a letter to the deputy head of each department "informing them of the results and identifying priorities to address in the coming year" (Price Waterhouse Coopers, 2009, p.17). The TSB initially assigned a rating to each area of management as well as an overall rating; however, such ratings are no longer assigned. The departmental reports are not made public but the Treasury Board does publish a government-wide report summarizing the results of the MAF across departments (Management Accountability Framework, n.d.).

The Federal Performance Project

The Federal Performance Project (FPP) was a collaborative effort between the Pew Charitable Trusts, the George Washington University, and *Government Executive* magazine. This instrument was created in 1998 as one of the first attempts to measure federal management performance for the purpose of assisting agencies to manage better. Additionally, the project intended to hold poorly-managed agencies accountable. At the same time, the idea was to help agencies learn from the management successes and failures of other agencies (Laurent, 1999).

This instrument collected information in five main areas:

- Financial management
- Capital management
- Human resources
- Information technology
- Managing for results

Grades were developed for the participating agencies in each of the five areas and were the result of a dialogue between academic personnel and journalists. The FPP was part of the larger Government Performance Project, which included state and local governments. The methodology for this instrument included a review of public documents, surveys, and interviews conducted by journalists from *Government Executive* magazine. The federal-level results were published in *Government Executive* magazine and received widespread attention. The state and local-level results were published in Governing magazine and were similarly the object of widespread attention (Ingraham et al., 2003).

OMB's Program Assessment Rating Tool

The Program Assessment Rating Tool (PART) was an assessment framework created by OMB during the Bush administration as part of the President's Management Agenda. The program was intended to inform decisions about how to allocate resources more efficiently and to hold managers accountable for the programs they implement. This instrument assessed programs across a set of performance-related criteria, including:

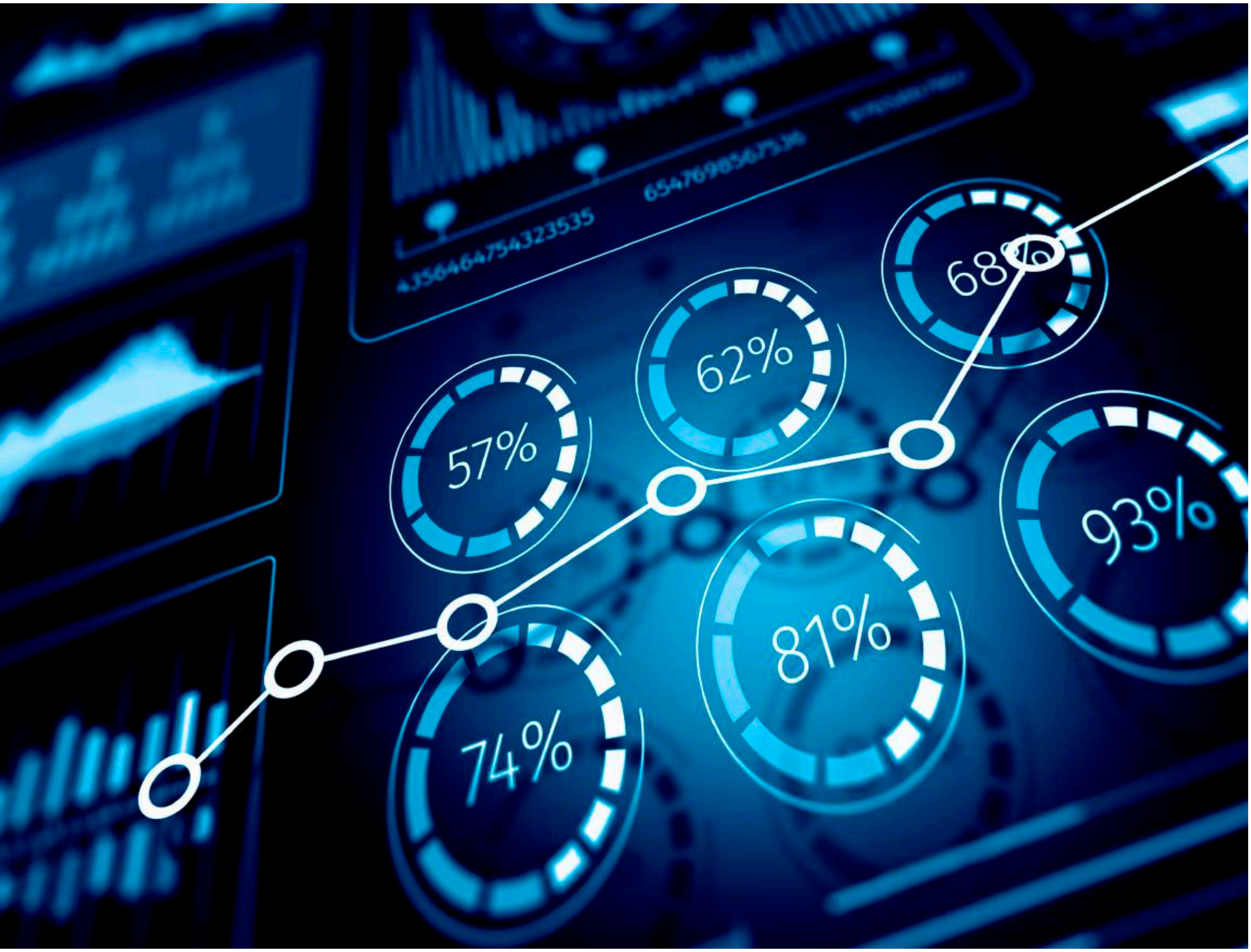
- Program design
- Strategic planning
- Program management
- Results

After the assessment, the programs were assigned numeric scores corresponding to qualitative ratings of:

- Effective
- Moderately effective
- Adequate
- Ineffective
- Results not demonstrated (U.S. Department of State, 2008)

The unit of analysis for the PART was the individual program rather than the agency as a whole although agencies were rated as red, yellow, or green in part on the basis of the PART outcomes consistent with the President's Management Agenda.

According to Gilmour (2006), the PART assessments were designed to have an impact in federal budget decisions; however, as of 2006 there was little evidence that PART had an impact on program management. Besides performance budgeting, the purpose of PART was to induce organizational change by encouraging agencies to find better ways of achieving their purposes. The PART measured outcomes instead of outputs, and a good grade in PART was intended to signify a well-managed program. For agencies and OMB, the challenge in implementing this program was to develop a good measure of performance and outcomes.



Appendix 3

Measuring the Management Quality of Federal Agencies: Workshop Participants

Wednesday, November 20, 2019

- Beth Angerman, General Services Administration
- Dustin Brown, Office of Management and Budget
- Ruth Buckley, U.S. Agency for International Development
- Dan Chenok, Stelic Center for the Business of Government
- Nani Coloretti, Urban Institute/Department of Housing and Urban Development (formerly)
- Lucy Cunningham, Partnership for Public Service
- Brodi Fontenot, Fontenot Strategic Consulting/Department of the Treasury (formerly)
- Robert Goldenkoff, Government Accountability Office
- Tim Gribben, Department of the Treasury
- John Kamensky, Stelic Center for the Business of Government
- Charles Keckler, Department of Health and Human Services
- Patrick Kennedy, Department of State (formerly)
- Emily Kornegay, Department of Housing and Urban Development
- John Koskinen, Department of the Treasury (formerly)
- Beth McGrath, Deloitte/Department of Defense (formerly)
- Alejandra Medina, University of Illinois—Chicago
- Gregory Parham, Department of Agriculture (formerly)
- John Rigg, Department of Health and Human Services
- Robert Seidner, Office of Management and Budget
- James Thompson, University of Illinois—Chicago
- Bill Valdez, Senior Executive's Association/Department of Energy (formerly)
- Danny Werfel, Boston Consulting Group/Department of the Treasury (formerly)
- Jim Williams, Schambach & Williams Consulting/General Services Administration (formerly)/Department of Homeland Security (formerly)

References

Baldrige Performance Excellence Program. (n.d.). Retrieved January 5, 2020, from https://www.nist.gov/baldrige.

Bloom, N., J. Van Reenen. 2007. Measuring and explaining management practices across firms and countries, *Quarterly Journal of Economics,* CXXII, (4), 1351-1408.

Bloom, N. and J. Van Reenen. 2010. Why do management practices differ across firms and countries? *Journal of Economic Perspectives,* 24(1): 203-224.

Congressional Research Service. 2009. *The President's Management Agenda: A Brief Introduction.* Washington, DC: Library of Congress.

Gilmour, J. 2006. *Implementing OMB's Program Assessment Rating Tool (PART): Meeting the Challenges of Integrating Budget and Performance.* Washington, DC: Stelic Center for the Business of Government.

Ingraham, P., P. Joyce, & A. Donahue. 2003. *Government Performance.* Why Management Matters. Baltimore, MD: John Hopkins University Press.

Laurent, A. 1999. Stacking up. *Government Executive.* February 1. Retrieved January 2020, from https://www.govexec.com/magazine/1999/02/stacking-up/6291/.

Management Accountability Framework. (n.d.) Retrieved January 2020, from https://www.canada.ca/en/treasury-board-secretariat/services/management-accountability-framework.html.

National Audit Office. 2009. *Assessment of the Capability Review Programme.* London: Cabinet Office.

National Academy of Public Administration. 2018. Strengthening Organizational Health and Performance in Government. Accompanying Working Paper #5. Case Studies of Other Countries' Use of Agency Capability Reviews. Washington, DC: National Academy of Public Administration

New Zealand State Services Commission. 2015. *Performance Improvement Framework: Core Guide 1: What is the Performance Improvement Framework?* Wellington, NZ: State Services Commission.

Panchamia, N. and P. Thomas. 2014. *Capability Reviews.* London: Institute for Government

Partnership for Public Service, Boston Consulting Group. 2019. *A Prescription for Better Performance: Engaging Employees at VA Medical Centers.* Washington, DC: Partnership for Public Service.

Performance.gov. (n.d.). 2019 revision to OMB Circular A-11, Part 6: Strengthening the Policy Framework for Improving Program and Service Deliver. Retrieved January 2020 from https://www.performance.gov/a-11-revision/.

Price Waterhouse Coopers. 2009. *Five Year Evaluation of the MAF.* Treasury Board of Canada Secretariat. Retrieved January 2020, from https://www.tbs-sct.gc.ca/maf-crg/implementation-implementation/fye-eq/fye-eq03-eng.asp.

Sadun, R., N. Bloom and J. Van Reenen. 2017. Why do we undervalue competent management? Neither great leadership nor brilliant strategy matters without operational excellence. *Harvard Business Review,* 95(5):120-127.

School of Government, Victoria University of Wellington. 2017. *Independent Review of the Performance Improvement Framework.* Wellington, NZ: Victoria University.

Sunningdale Institute. 2007. *Take-off or Tail-off? An Evaluation of the Capability Reviews Programme.* Ascot, England: National School of Government.

The Drucker Institute. (n.d.). Company rankings. Retrieved January 2020, from https://www.drucker.institute/programs/company-rankings/.

Treasury Board Secretariat. 2017. Evaluation of the Management Accountability Framework. Retrieved January 2020 from https://www.canada.ca/en/treasury-board-secretariat/corporate/reports/evaluation-management-accountability-framework.html.

U.S. Government Accountability Office. 2015. *Managing for Results: Practices for Effective Agency Strategic Reviews.* Washington, DC: Government Printing Office.

U.S. Government Accountability Office. 2017. *Open Innovation: Executive Branch Developed Resources to Support Implementation, but Guidance Could Better Reflect Leading Practices.* Washington, DC: Government Printing Office.

U.S. Office of Personnel Management. 2018. *Human Capital Reviews.* Washington, DC: Office of Personnel Management.

U.S. Department of State. 2008. *FY 2008 Performance Summary.* Retrieved January 2020 from https://2009-2017.state.gov/documents/organization/79676.pdf.

World Management Survey (n.d.). Retrieved January 2020 from https://worldmanagementsurvey.org/.

Get in Touch

1725 I Street, NW
Suite 300
Washington, DC 20006
Phone: (800) 913-1883
Email: stelic@stelic.com

Amelia Cruz, Director of Consulting
Phone: (800) 550-4966 ext 501
Email: acruz@stelic.com