

**Change and Continuity: Lessons Learned from the Bush and Obama Administrations’
Experiences with Evaluation and Performance Measurement**

Nicholas R. Hart

The Trachtenberg School of Public Policy and Public Administration
The George Washington University
Washington, DC, USA

Kathryn Newcomer

The Trachtenberg School of Public Policy and Public Administration
The George Washington University
Washington, DC, USA

Paper to be presented at the

American Evaluation Association Conference

Evaluation 2015 "Exemplary Evaluations in a Multicultural World"

Chicago, IL, USA

November 5, 2015

DISCLAIMER: The views presented in this paper are those of the authors and do not represent the official position of the U.S. Government or the U.S. Office of Management and Budget (OMB).

Abstract

Calls for "evidence-based policy" and for assessing how well government programs work have been around for many years. The George W. Bush and Barack Obama Administrations both espoused support for the generation and use of evidence to guide and improve government management. The two presidents brought very different professional experiences, political views, and policy advisors to the job as Chief Executive of the federal bureaucracy, yet their "President's Management Agendas" established similar expectations about the use of evaluation and performance data. The paper outlines how the two Presidential Administrations centrally approached "evidence-based policy" and "performance management," with emphases on program evaluation and performance measurement, respectively. We highlight the many similarities across the Administrations, the interesting differences, and the intriguing ways in which some lessons that could have been "learned" were not. The role of the Office of Management and Budget (OMB) and pertinent laws are detailed, and the experience of the Federal executive agencies implementing the Administrations' agendas are summarized.

**Change and Continuity: Lessons Learned from Two Administrations’
Experiences with Evaluation and Performance Measurement**

Over the past quarter century, increased calls for government accountability, learning, and improved program performance in the United States transformed how the Federal government implemented so-called “performance management” systems. These systems generally emphasize the acts of goal-setting, performance measurement, and related reporting. More recently, some efforts to promote "evidence-based policy" sought to steer resources based on the results of impact evaluations, which seek to estimate whether an intervention at a particular time and place caused a change in a desired outcome. The first two presidents of the 21st Century—George W. Bush and Barack Obama—both led administrations that encouraged these activities, among others. Despite different policy priorities, ideological perspectives, professional backgrounds, and senior advisors, both Bush and Obama established performance and management agendas interconnected in their aim to promote government performance.

In this paper we suggest that, notwithstanding the notable differences between the Bush and Obama performance agendas, the two administrations' plans featured a strikingly large number of similarities. The similitude may have contributed to strengthening overall implementation of reform efforts and paved a path for even more transformative efforts in the future. Yet, the similarities and differences between approaches pursued by each respective administration highlight interesting opportunities where lessons about implementing these reforms could have been perceived and learned by officials centrally charged with implementation, but do not appear to have resulted in learning. While performance measurement can be thought of as a subset of evaluation, or at least a complement, in the sections that follow,

evaluation and performance monitoring activities are generally discussed separately, consistent with the manner in which the activities are implemented in the majority of Federal agencies and programs.

The paper will first proceed by considering what evidence-based policy is generally intended to be and the overarching goal that some advocates of the strategy articulate. The paper will then proceed by separately presenting an overview of the Bush and Obama initiatives, accounting for lessons learned in each administration and distinguish where efforts diverged and aligned. The paper will then consider several areas where lessons could still be learned to inform future efforts and conclude with a perspective on addressing integration of performance measurement and evaluation activities within the Federal government.

I. Transitioning From "Results-Driven" Toward "Evidence-Based" Policymaking

Following the inception of new, major social investments during the mid-20th century, the field of public administration shifted increasingly toward democratic management practices and saw the advent of the New Public Management (NPM) school. The NPM philosophy evolved from a belief that government failures, like market failures, can be catastrophic for productivity and livelihoods (Moe, 1987). Federal budgeting shifted toward attempts to implement a Planning-Programming-Budgeting System and zero-based budgeting in the 1960s and 1970s, with emphasis on applying business-like principles to running government, to the point such attempts became both widespread and common (Barzelay, 2001).

By the 1990s, Osborne and Gaebler's (1992) seminal work on government reform advocated a series of principles for improving public programs based on select case studies in implementing innovative government interventions, many of which served as the core principles for the Clinton Administration's National Performance Review (NPR). NPR was explicitly

focused on ways to "create a government that works better and costs less" by targeting efforts to prioritize customer service, reduce bureaucratic hurdles, empower federal employees to innovate with processes, and to refocus on agency mission critical activities (Gore, 1998). The underlying philosophy was to take inefficient government programs and transform them in to more effective, more efficient, and more dynamic operations. While the evidence-based policy movement is rooted in efforts that preceded NPM, the NPM school helped increase the demand for evaluation and analysis to guide decision-making on government policies (Dahler-Larsen, 2011). This idea of identifying and addressing the inefficiencies and ineffective activities with some level of analysis serves as the core tenet of evidence-based policy, as discussed by proponents in recent years.

By the mid- to late-1990s, the Government Performance and Results Act of 1993 (GPRA) began to fundamentally change the way federal programs reported to Congress, the White House, and the public on program performance, shifting from an emphasis on program inputs to instead focus on program outputs and outcomes. Once fully implemented beginning in fiscal year 1999, GPRA required agencies to improve planning through periodic updates of strategic plans and the development of performance-based budgets, where each year the annual budget submission to Congress would include key metrics for programs. GPRA's performance measurement requirements were premised on the idea that tracking performance would help agency leaders, Congress, and the public to hold programs accountable for achieving program objectives, continuous improvement, and using available performance information to inform subsequent decision-making, particularly for the budget process (Joyce, 2011).

Osborne and Gaebler (1992) and the Clinton administration's NPR advocated a results-driven mantra, incorporating a process by which output and outcome information contributed to

program management to provide feedback on operations and the extent to which program objectives are met. The NPM school places these discussions in the context of limited or declining resources, which in turn necessitates program efficiency and prudent use of taxpayer resources. A newer concept of evidence-based decision-making and practices, which aligns with the rational choice model, places greater emphasis on “rigorous” program evaluation, which it defines as primarily involving randomized controlled trials (RCTs), and implicitly discounts the value of many other evaluation methods, including performance measurement.

While advocates for this conception of evidence-based policy highlight goals similar to results-driven proponents—the potential gains to be made in the number of individuals served and alternative uses of limited budgetary resources—these evidence-based advocates also suggest programs can be independently prioritized and even expanded to broader scale based on identified successes from RCTs. For example, former director of the U.S. Office of Management and Budget (OMB), Peter Orszag, emphasized that “By instilling a culture of learning into federal programs, we can build knowledge so that spending decisions are based not only on good intentions, but also on strong evidence that carefully targeted investments will produce results” (Orszag, 2009). So while these evidence-based decision-making advocates similarly support improvements in program efficiency, their long-term strategy might be best articulated as using RCTs to build evidence as a means for reducing government funding for ineffective programs and re-directing funding to interventions that have shown an impact in an RCT.

II. Bush Administration Performance and Evaluation Reforms

The Clinton NPR initiative evolved over the course of the 1990s, and by the end of the Administration had been rebranded as the National Partnership to Reinvent Government. The modified initiative took aim at narrowly targeted agencies instead of the entire Federal

government, namely focusing on agencies that provided direct customer services to the public. Specific areas of management focus became better coordinating services related to electronic government, human capital, and customer service. By certain accounts NPR efforts succeeding in achieving some discrete goals and maintaining focus on the initiative (Gore, 1997), but by other accounts NPR did not fully engaged political leadership, improve trust in government programs, integrate the initiative with performance measurement, or consult with congressional stakeholders (Kettl, 1998).

President's Management Agenda (PMA)

Building on the efforts of the Clinton Administration's NPR initiative and other earlier efforts, shortly after assuming office in 2001, the Bush Administration launched a new President's Management Agenda (PMA). The Bush PMA focused on reforms for discrete mission-support areas such as managing human capital, improving financial performance, contracting to promote competition, improving information technology, and budget and performance integration, in addition to a series of agency-specific goals (OMB, 2001). Each agency was scored quarterly by OMB staff based on their performance and successes in meeting centrally-defined targets for each reform area. To help coordinate the PMA, President Bush re-established the President's Management Council (PMC) which included agency Chief Operating Officers and served as a mechanism to oversee implementation of the PMA (U.S., 2008b).

One feature common to multiple elements of the PMA priority reform areas was the use of scorecards, or short summaries of how agencies fared in meeting specified objectives, typically reflected with traffic light indicators. The scorecards were identified as useful by some agency staff who believed the tool could help draw policy officials' attention to important issues (GAO, 2006). OMB updated and publicly released the scorecards each quarter, denoting the

status of achieving the goals as unsuccessful, mixed, or successful, and whether agencies were making progress in accordance with timelines established by each agency.

Program Assessment Rating Tool (PART)

Perhaps the Bush Administration's performance legacy will be best remembered for the creation, implementation, and dissemination of the PART. The fiscal year 2004 budget offered a critique of GPRA that the statute only required the collection of performance data, but no formalized process for using those data (OMB, 2003). The Bush Administration designed the PART process to address that issue and to do so in a public and transparent way. Designed as an accountability tool, PART was intended to help fulfill the PMA's budget and performance integration reform goal by better linking performance information in budget decision processes.

PART was designed as a 25 question instrument with emphasis on program purpose and design, strategic planning, program management, and program results. Each section included specific questions, with differential weighting, used to produce an overall score for a program. Half of the total score was based on whether programs demonstrated progress in achieving long-term outcomes, met annual performance and efficiency goals, and whether the program was supported by "independent" evaluations. The initial instrument was developed in partnership with Federal agencies, with feedback incorporated from 67 programs who pilot tested the assessment tool (OMB, 2003).

The process of rating each program began with a negotiation between OMB and agencies on what constituted a program for purposes of the PART (Joyce, 2011). Programs could be defined based on a variety of approaches, including appropriations line items or other levels agreed to by agencies and OMB. Central coordinating staff, typically in agency budget offices, worked with program staff to conduct self-assessments for each question in the PART

instrument. Agencies prepared supporting documentation and submitted the self-assessment to OMB for review along with volumes of supporting documentation when available for each program. OMB career program examiners reviewed the documentation to finalize the assessment. In cases where OMB staff did not agree with the agency's assessment, OMB's rating could supersede the agency but typically only after discussion among senior policy officials in OMB and the respective agency, and in later years appeals of OMB decisions were made directly to a panel of PMC members (OMB, 2008).

Once a program received a PART composite index and rating, scores were made publicly available on ExpectMore.gov (U.S., 2008a). This was the first time a major government reform effort disclosed information about specific programs, using a broad definition of the term program. Clinton's NPR did not release all of the drafted reports (e.g., Executive Office of the President). In addition to the release of PART scores, public summaries were provided, including a one-page cover note and a detailed description of how programs fared on each of the questions in the composite score.

PART was never touted by the Administration as a perfect tool for assessing program effectiveness or utility. Instead, the Bush Administration routinely solicited feedback to improve the value and utility of the PART. OMB staff who developed the PART process incorporated feedback from agencies, and even considered public and other external stakeholder comments in the design (OMB, 2003). In some cases, questions viewed by agencies as overly subjective were eliminated. For example, the question "Is the Federal role critical?" was removed from the PART instrument based on feedback from agencies (OMB, 2003). GAO (2004) identified a suite of more than 30 modifications in the instrument between the 2004 and 2005 budgets, including wording clarifications, additional questions, dropped questions, and merged questions.

Within OMB, while individual program examiners in Resource Management Offices were charged with conducting the reviews, the PART design and consolidation was centrally managed by OMB's Performance Evaluation Team (PET), a group of OMB examiners from throughout the agency, and later the Division of Performance and Personnel Management. This office of primarily OMB career staff was charged with issuing government-wide guidance, providing technical assistance to agency and OMB staff conducting assessments, developing the public website, ensuring assessments were completed within given timeframes, and coordinating responses to agency appeals. The ability of a central coordinating team to articulate timelines and deadlines to OMB program examiners, who had PART responsibilities layered on top of existing responsibilities, helped keep the production timelines on schedule for PART reviews. The central coordinating team served the role of reducing complexity for internal guidance provided to OMB program examiners to simplify and streamline their process decisions amidst competing priorities and other demands on their time.

Over the course of the Bush Administration, over 1,000 programs received PART scores, including hundreds with multiple iterations of the assessment updated with new information over time (U.S., 2008a). By the end of the Administration, the PART process resulted in Federal government having more than 6,000 program performance metrics and more than 1,000 efficiency measures, many of which were created as a result of the PART process. But the proliferation of performance measurement did not necessarily result in more evidence-based policymaking. Among the thousands of performance measures created during the Bush Administration were hundreds (if not thousands) that presented the facade of managing by results when the measures were instead created merely as a compliance exercise (Moynihan, 2013). For instance, within the newly formed Department of Homeland Security the United

States Secret Service proposed outcome goals to achieve zero assassinations and injuries of protectees (U.S., 2008a). While such goals are societally desirable, the existence of such goals in driving actual agent performance is of questionable utility.

Performance Improvement Council

During President Bush's second term, the president signed an executive order creating the Performance Improvement Council (PIC). The PIC was envisioned as a coordinating body to continue the same reform objective of the PMA that the PART addressed, but by identifying senior leaders in Federal agencies to coordinate performance management activities. Specifically, PIC officers were directed to oversee updates to agency strategic plans, ensure that aggressive program goals were set with reliable measurement procedures, routinely convene program staff to assess performance, and work to incorporate performance measures in personnel appraisals (OMB, 2007). Many agencies ultimately nominated their chief financial officers for this role. The PIC formally convened in January 2008 and met monthly through the end of Bush's term, with subcommittees focused on discussing issues related to program evaluation, best practices, and transparency.

The Bush Administration also envisioned institutionalizing performance reviews through a decennial bipartisan commission, charged with convening periodically to consider ways to improve the performance of Federal programs, and even recommend terminating sacrosanct but ineffective programs (Breul, 2007). While the proposed legislation was not enacted by Congress, it served as a benchmark in the results-driven policy discussions focused on the need for programs to continually seek performance improvements or be at risk of increased scrutiny or termination in future discussions.

Program Evaluation

The 70-page Bush PMA work plan produced by OMB included just one bullet focused explicitly on using program evaluation. The bullet, entitled "demand evidence" noted the perceived lack of rigorous evaluations for many programs and stated that rigorous evidence should be requisite for Federal funding (OMB, 2001). Much of the Bush Administration's emphasis on program evaluation was conducted through the PART process, described above, which included two questions on the topic of evaluation. The two questions included in the PART instrument focused on summative or impact evaluation and generated extensive feedback from external stakeholders and agency staff on the topic of what constitutes "strong" evidence.

In 2004, OMB issued a guidance document entitled "What Constitutes Strong Evidence of Program Effectiveness" to help agency staff and OMB program examiners in preparing PART ratings (OMB, 2004). The guidance, which was drafted in concert with the Coalition for Evidence-Based Policy, an advocacy group, emphasized RCTs as the premiere method for impact evaluation, with the majority of the document articulating the benefits of RCTs relative to other forms of evaluation. The guidance did note that RCTs are not appropriate in all cases, but generally discounted the value of other evaluation methods for assessing impact. In addition, some interest groups and good-government advocates promoted RCTs as the only proper form of evaluation in training sessions held for agency and OMB staff (Haskins & Margolis, 2014). In response to this formalized, narrow perspective on high quality evaluation, professional evaluators from federal agencies sought to influence OMB regarding when other evaluation could be appropriate, in response to the PART's emphasis on impact evaluation, and engaged in a dialogue with OMB (Bernholz et al., 2006). In addition, some professional evaluators requested that the administration provide more context on the use of RCTs, identify limitations

regarding construct and external validity, and stress the importance of mixed methods, among other comments (Trochim, 2008).

The Bush evaluation agenda embedded within PART also incited extensive discussion about what constitutes an independent evaluation. The initial pilot and fiscal year 2004 rounds of PART only considered evaluations independent if they were conducted by inspectors general, the General Accountability Office (GAO), or external evaluators with non-governmental funding. In subsequent years, OMB broadened the definition of independent evaluation to include agency evaluation offices and third party contracted evaluations (GAO, 2004). Notwithstanding the feedback from professional evaluators on the issues of rigor and independence, recent work to understand PART scores suggests programs that relied on internal evidence or qualitative evaluations tended to have lower composite PART ratings (Heinrich, 2012), potentially suggesting methodological bias among OMB examiners.

Other efforts, exclusive of PART, to pursue evaluation in the Bush Administration were met with mixed results. While programs with statutory evaluation directives and mandatory funding streams continued to produce evaluations, including those with modifications from the Deficit Reduction Act (DRA) of 2005 (e.g., the Department of Health and Human Services Administration for Children and Families), new evaluations sought by the Administration faced some opposition in Congress. For example, the Bush Administration's plan to conduct a new RCT experiment for the Upward Bound program at the Department of Education was ultimately barred by law in the fiscal year 2008 appropriations bill ("Consolidated Appropriations Act, 2008," 2008).

Finally, in an effort to improve the dissemination of evaluation results, some agencies initiated plans to provide public portals for evaluations sponsored or funded by an agency.

While some programs, such as HHS's Administration for Children and Families had released evaluation results for years, additional agencies saw value in better disseminating findings and followed suit. For example, in 2002 the Department of Education's Institute for Education Sciences launched the What Works Clearinghouse (IES, 2014) and in 2006, the Substance Abuse and Mental Health Services Administration launched a comprehensive website called the National Registry of Evidence-Based Programs and Practices (SAMHSA, 2014).

III. Lessons from the Bush Administration Performance Management Reforms

Reviews of the Bush Administration's performance and evaluation strategies are now widely available, highlighting positive aspects, along with some of the more limiting features of the strategies (Heinrich, 2012; Joyce, 2011; Moynihan, 2013; Moynihan & Lavertu, 2012; White, 2012). Regardless of the level of success, implementation of the PMA and the Bush Administration's perspective on evaluation offer potential lessons for future performance improvement and evidence initiatives.

- **Purpose.** The Bush Administration's PART instrument sought to fulfill multiple purposes – providing information for public and congressional accountability as well as informing congressional and executive funding decisions (Newcomer & Redburn, 2008). The attempt to serve many purposes and a wide range of constituents limited the Administration's ability to effectively target the tool in a manner that could support both program improvement goals and accountability objectives, particularly when the two goals conflict. Metzenbaum (2009) suggested that presenting information to serve specific audiences should be a guiding principle for future reform efforts. Newcomer and Redburn (2008) similarly identified that it is difficult to design efforts that can both

simultaneously track performance and assess effectiveness, and the trade-offs should be weighed carefully.

- **Accountability and Transparency.** PART was designed to publicly render judgment on Federal programs with budgetary consequences, since the performance monitoring required by GPRA alone had produced little results (Breul, 2007). The assignment of PART scores did enable public transparency of Federal programs in a way that had never been done, but the provision of single composite scores, and the categorical names for the ratings applied (e.g., “ineffective”) likely contributed to the apprehension among agency staff for supporting the process. In particular, the presumption that a single tool could guide funding decisions must account for the reality that performance information is subjective in nature, and an indication of weaker than expected performance might produce different normative perspectives on the diverging needs to either increase or decrease a program's funding. Thus, performance information is one of many factors decision-makers address when determining an appropriate level of funding for a program.
- **External Stakeholder Support.** While Federal agencies expended considerable resources conducting PART reviews and implementing other PMA reforms during the Bush Administration, some observers have argued that Congressional interest remained minimal (Breul & Kamensky, 2008). Given the broader policy context and a strong economy for the majority of the Bush term, efforts to identify a mechanism to support and inform funding decisions may not have gained traction because it was not perceived as necessary. Further, given limited congressional interest, agency participation may have been dis-incentivized for senior officials since it did not appear likely the PMA

reports would affect agency budgets approved by Congress (Newcomer & Redburn, 2008).

- **Senior Official Engagement.** The Bush Administration used the PMC and PIC to meaningfully engage senior policy officials from agencies to focus on performance management, which impelled some level of participation from agencies, even if minimal. In addition, some performance improvement officers started to play an important role in reviewing and coordinating performance measurement activities near the end of the second Bush term. By setting an expectation that performance is a priority and that agency heads and policy officials are accountable for demonstrating agency results, senior executives may have been more likely to engage in discussions on overall agency performance.
- **Agency Participation and Staff Workload.** One challenging aspect of the PMA activities was that the reforms required considerable resources from both agency staff and OMB (Moynihan, 2010). Completing a PART review with full documentation and briefings for appropriate agency leaders and OMB typically required a dedicated coordinating staff. In addition, OMB program examiners had PART as well as other PMA activities layered on top of their normal budget and policy responsibilities. Both agencies and OMB experienced limits on the capacity of the staff to provide enough attention to adequately produce PART reviews, which distracted from other tasks (Joyce, 2011; Newcomer & Redburn, 2008; White, 2012). Given the high degree of workload associated with PMA activities, agencies may have come to view the requirements as acts of compliance rather than performance enhancements (Newcomer & Redburn, 2008). To the extent agencies viewed participation as a compliance exercise, early engagement with

agency staff to solicit comment on the design and use of the PMA scorecards and the PMA instrument likely eased tension with some agencies.

- **Uniformity-Flexibility Trade-off.** The PMA scorecards and the PART provided standardized reporting templates for agencies to complete, without fully recognizing the differences in agency activities, functions, and institutional structures. While the uniformity facilitated central management within OMB, agencies had less flexibility to address the PMA criteria in a manner appropriate for individual agency operations. The focus on uniformity may have distracted from efforts to make sense of performance trends and the magnitude of problems faced to achieve meaningful improvements in program performance (Metzenbaum, 2009). Despite a quest for uniformity, there were agency staff perceptions of subjective and unfair treatment in assessing program performance by OMB examiners.
- **Devolution of Goal-Setting.** One trade-off made during the PMA's PART implementation was shifting goal setting to the program level, in addition to allowing agency policy officials to establish agency-wide objectives under GPRA. The focus in the Bush Administration was on setting and achieving goals, but little emphasis was placed on innovation, nor on discovering better ways to run public programs (Metzenbaum, 2009).
- **Staff Training.** For both evaluation and performance measurement activities, new processes require additional training for agency and OMB staff to ensure a common understanding of implementation goals and how to address challenges. The Bush Administration's coordinating team within OMB engaged agency staff to some extent, and offered training to OMB program examiners on the principles of performance

measurement development and generally on program evaluation, with a primary focus on experimental methods.

- **Performance and Evaluation Integration.** The fact that the PART instrument posed questions about the presence of evaluation was a small step toward better integrating evaluation and performance monitoring activities. The PART question, in the framework of the structured performance tool, demonstrated the importance of evaluation to the Administration and a recognition that evaluation alone cannot improve program performance. The incorporation of evaluation questions in the PART set the stage for OMB to frame the use of evaluation as a learning and program improvement tool that could be used in conjunction with other performance management resources.
- **Definition of Evidence Quality.** The Bush Administration's sidebars on evaluation independence and focus on RCTs to provide methodological rigor suggested a disconnect between administration officials interested in better pursuing evidence-based policy and a broad cross-section of program evaluation practitioners and theorists. Working with a broader range of agency staff, social scientists, and professional evaluators in advance of issuing program guidance on evaluation requirements may help provide a common understanding of what is meant by evidence and how the administration may most credibly achieve that goal. The Bush Administration did consult with evaluators, but only after OMB's narrowly defined guidance met with some objection by the professional evaluation community.
- **OMB's Strategic Role.** Throughout the implementation of the PMA, the Bush Administration used OMB staff to centrally coordinate activities. OMB staff, who were instrumental in delivering the PART instrument to agencies and subsequently

coordinating and approving the assessments, were situated well to understand the prioritization of the PMA to the Administration and familiar with agency staff charged with developing responses to the various directives. While individual OMB program examiners may have shown some inconsistencies in PART scores (Moynihan, 2013), unreasonableness in calling for more efficiency measures (Metzenbaum, 2009), or possessed insufficient knowledge of certain programs at the initiative's outset (White, 2012), PART would not likely have been implemented at all if not for OMB's coordinating role. At the same time, PART and the other PMA reforms provided OMB examiners with valuable leverage in engaging with agency counterparts, to formally discuss program performance, create new performance measures where they may not have existed before, and to encourage the production of new program evaluations.

- **Implementation Planning.** Implementation of priority initiatives may often be an afterthought for administrations seeking to announce new reforms as quickly as possible. Yet, identifying a clear implementation process at the outset can help ensure reforms are appropriately scaled and time to meet an administration's needs. When implementing the PART, OMB engaged agencies for feedback before fully launching an effort to formally discuss agency performance. Then, the burden of proof for justifying the PART self-assessment was placed on agency staff who were obligated to at least defend performance goals well enough to satisfy OMB staff (Moynihan, 2013). While the results of the process may not have been accepted by all, the process for implementing PART and the other PMA reforms was clearly articulated through dozens of guidance documents for agency staff to provide for implementation of the initiative.

The Bush Administration's experience with the suite of PMA reforms, including PART, provides numerous lessons to inform future administration's efforts, as discussed above. In the final year of the Bush Administration, multiple analyses emerged with findings and recommendations for the next administration (Metzenbaum, 2009; Newcomer & Redburn, 2008), some of which were considered in designing the Obama Administration's initiatives, while others were instead re-learned.

IV. The Obama Administration's Performance and Evaluation Reforms

The Obama Administration's perspective on evaluation and performance measurement evolved over time, framed by different political contexts in which the Administration has operated. Even before President Obama assumed office in 2009, his campaign strategy reflected a strong reliance on data analysis. The election campaign employed a strategy that utilized small-scale RCTs to target effectively potential voters, campaign donors, and volunteers (Issenberg, 2012a, 2012b; Showalter, 2013). In one example, call centers were used to assess which voters were likely to be persuaded by certain campaign messages, which resulted in the Obama campaign's "voter persuasion" model (Issenberg, 2012a). The early use of RCTs in the Obama campaign strategy seemingly helped set the stage for future discussions about the use of data analytics in the Administration.

The Obama Administration's early espoused interest in performance and evaluation was rooted in trying to identify a "common sense" approach for increasing government effectiveness (U.S., 2015). The President said "we know there's nothing Democratic or Republican about just doing what works. So we want to cast aside worn ideological debates and focus on what really helps people in their daily lives" (Obama, 2009). The strategies employed by the Obama Administration fit in to four categories: (1) developing an overarching performance management

framework, (2) replacing PART by instituting priority goals, (3) facilitating the production and use of impact evaluations for government programs, and (4) encouraging the use of third party tools to support innovative projects that satisfy government objectives.

Management Agenda

The Obama Administration's initial performance management agenda, similar to the Bush Administration, focused efforts on addressing issues related to information technology, government real property, human capital, and customer service. The initiative was described as focusing on four pillars: effectiveness, efficiency, economic growth, and people and culture (OMB, 2015). Of note from this agenda is the emergence of a series of Stat processes created to review data and to identify paths forward in improving efficiencies. These reviews included PortfolioStat for IT investments, PerformanceStat for performance information, and HRStat for human capital. In 2015 the Stat model began to transition with the ultimate goal of conducting one collective annual review of the various management agenda items, including the priority goals discussed below.

Priority Goals

The fiscal year 2010 budget, published just months into the Obama Administration's first term, did not include the PART. While recognizing PART's value for establishing new performance measures and targets across the Federal government, the Obama Administration viewed it as less successful at achieving actual improvements in performance for the President's priority areas (OMB, 2009a). Much like the critiques the Bush Administration offered that GPRA only resulted in the creation of performance information, the Obama Administration viewed earlier reforms as unsuccessful in facilitating widespread use of performance information (OMB, 2010b).

The Obama Administration's approach to improve performance was to create high-level priority goals within and across agencies to help achieve tangible improvements in program performance (Mark & Pfeiffer, 2011; OMB, 2009a). The new performance plan was to be led by a new Federal government executive position called the Chief Performance Officer, and Federal agencies would be required to publicly demonstrate progress toward the respective goals, to explain progress, and to hold senior managers and leaders accountable for goal attainment (Brass, 2011; OMB, 2009a).

High Priority Performance Goals. In mid-2009, OMB outlined a process for developing near-term high priority performance goals (HPPGs) for major agencies in the following year's budget (OMB, 2009e), that would ultimately be incorporated in to annual performance and strategic plans. OMB focused on improving the use, communication, and networking for the proposed performance system (OMB, 2010a, 2010c). OMB directed agencies to focus on priority goals with "unrelenting attention," develop review processes to ensure sustained agency performance, and work toward making GPRA documents more useful. OMB suggested that HPPGs should reflect direct public value or key agency missions, have existing authorizations and appropriations, represent areas with implementation challenges that could improve effectiveness if resolved, have evaluable outcomes, and address challenges that would not likely be overcome in the absence of targeted agency resources.

HPPGs were focused on articulating agency expectations to achieve results, similar to other GPRA efforts, and were intended to target key objectives in which an agency's senior leaders were interested and where agencies could commit resources. Processes for HPPG development and target setting were to be decided by agency leadership, largely without the input of external stakeholders in Congress or OMB, which played a central role in performance

measurement construction during the Bush Administration. Agencies were provided considerable flexibility and autonomy in specifying priorities, goals, and targets through an agency-driven process (Joyce, 2011). And with that came the responsibility of assigning senior agency leaders who were expected to review and report quarterly on goal progress. The design of HPPGs seemed to incorporate a lesson from the Bush PMA and other efforts to integrate performance and budget information, that there are limits to the amount of information that can be centrally coordinated and processed for decision-makers in monitoring overall agency performance (i.e. bounded rationality).

Similar to performance measurement transparency and accountability aims of PART, OMB planned to disseminate priority goals through a single Federal performance portal on which performance could be tracked and publicly available. Once the priority goals were established, the portal would include them and all GPRA performance information (OMB, 2010c). When the goals were released, the Obama Administration initially established a list of 126 priority goals across the largest Federal agencies. For example, the Department of Housing and Urban Development and the Department of Veterans Affairs jointly announced an ambitious goal to address veterans' homelessness, reducing from a baseline of about 200,000 homeless veterans to just 59,000 within the two year window (OMB, 2011c). Measured in 2014, the Administration reported that less than 50,000 veterans were homeless, which resulted in a second iteration of the goal being modified to end veterans homelessness altogether (U.S., 2015).

GPRA Modernization. In early 2011, Congress enacted the GPRA Modernization Act. The legislation provided four key changes to GPRA and codified much of the existing practice from the Obama Administration, with slight modifications. The GPRA Modernization Act (1) established and defined roles and responsibilities for agency heads, chief operating officers,

performance improvement officers, and agency goal leaders, (2) mandated establishment of priority goals, including long-term cross-agency goals set by OMB and agency priority goals updated by agencies on a biennial basis, (3) required periodic reviews of goals, and (4) modernized reporting mechanisms through the creation of a single website (OMB, 2011b). One of the new features of GPRA Modernization was aligning the timeframes for strategic planning, performance reports, and goal setting requirements with the election cycle and President's annual budget submission, effectively allowing plans and goals to better align with the President's priorities (Brass, 2012).

In implementing GPRA Modernization, OMB outlined the objectives of the new performance management framework – effectiveness, efficiency and productivity, transparency that engages the public, and fairness and equity (OMB, 2011a). Similar to the Administration's original HPPG plan, the new agency priority goals did not require consultation with Congress or other stakeholders, in conflict with the transparency and collaboration objectives outlined by the Administration. To target resources to the highest priorities, OMB applied priority goal requirements only to the 24 agencies covered by the Chief Financial Officers Act of 1990, a flexibility provided to the executive branch by the GPRA Modernization Act (Brass, 2012; OMB, 2011a, 2011b). OMB further directed that *new* agency priority goals be achievable within 24 months, solidifying their short-term design, and clarified that the HPPGs previously developed would be "archived" (OMB, 2011b).

OMB's role in setting and reviewing targets for the agency priority goals was limited, just as with the HPPGs. By limiting this role, the Administration may have ensured agencies were setting goals for existing activities or areas that would have been accomplished regardless of the initiatives. Under the statute, agencies are responsible for reviewing their progress on the various

goals quarterly. However, for two other types of goals—cross agency priority goals (CAP) and management goals—OMB is charged with setting the annual and quarterly targets, as well as convening relevant agencies quarterly to review progress (OMB, 2011b).

The evolution of the priority goal process over the course of the Obama Administration reflects a desire from initiative designers to allow for wholesale change in priorities and performance goals over time. For example, the public website disclosing the priority goals, performance.gov, was updated in 2014 to reflect only the most recent cohort of priority goals, archiving all goals and performance information from 2012 or earlier (U.S., 2015). While the public website does note progress toward achieving each respective goal, the Administration avoided aggregating the number of goals on track or behind schedule to succinctly summarize the initiatives collective progress. In October 2015, Administration officials announced the final set of 92 new APGs for Federal agencies during the second Obama term. The announcement emphasized the Administration's progress in reducing veterans homelessness, attributed to an early priority goal established to facilitate some collaboration across agencies (Donovan, 2015).

During the first Obama term, priority goals received intense focus from the Administration, while in the second term as the goal production processes became more institutionalized, new guidance on the goals became less frequent. Guidance focused on continuing to implement the goals based on near-term deliverables, even as the government engaged in a process to set new agency and cross-agency goals (OMB, 2013b). More recently, the Administration has suggested implementation of the goals has been successful, but substantial challenges exist to coordinating improvements that span across multiple agencies (OMB, 2015).

Strategic Reviews. One requirement of the GPRA Modernization Act was for agencies to annually review progress toward achieving strategic goals, with the goal of incorporating that information into budget requests, to focus long-term strategies, and in some cases to manage program risk. OMB's guidance on the reviews indicates they are primarily agency led, although OMB maintains a role in suggesting process improvements, helping identify budget and legislative options to address major issues, and reviewing public performance plans and reports to ensure planned actions align with the President's policies (OMB, 2014). Initial results of the strategic reviews and the progress on strategic agency goals were first publicly reported in February 2015 for about three-quarters of the participating agencies, although the Obama Administration did not aggregate the results to publicly indicate the extent to which agencies are demonstrating progress (U.S., 2015).

Evaluation Capacity Building

Obama Administration officials declared a prioritization of program evaluation activities early in the president's first term, although avoided describing an integration with the priority goal efforts (OMB, 2009a, 2009b, 2009e). In late 2009, OMB articulated the Administration's evaluation philosophy by specifying that the Administration supported "*rigorous, independent program evaluations*" (OMB, 2009b). Over the course of the administration the description of evaluation's value would increasingly focus on an ability to "find and fix or eliminate ineffective [programs]" (OMB, 2011e). The acknowledgement of fixing programs focused the discussion on using evaluation to assess causality for determining whether to fund or not fund programs, aligning with the NPM school and similar rationales that manifested during the Clinton Administration.

Much like the evaluation philosophy during the Bush Administration, the Obama Administration focused primarily on increasing the use of *impact* evaluation, including justifications that existing evaluations had insufficiently influenced budget and management decisions, and that funds had been spent on evaluations of insufficient rigor or policy insignificance (OMB, 2009b). The evaluation agenda also prioritized shifting investments to "what works and less [to] what does not" through summative evaluations aimed at identifying causal effects (OMB, 2009b, p. 1). While OMB's guidance did not explicitly identify the potential benefits of formative evaluation for organizational learning, it did acknowledge a range of methods and the importance of selecting the most appropriate methods for addressing evaluation questions, which was widely supported within the evaluation community (Rog, Trochim, & Cooksy, 2009a). Yet, OMB did not specifically acknowledge qualitative methods until well in to the Obama term, citing the methods as a useful technique to understand additional program context (OMB, 2011e). By the second Obama term, senior policy officials from across the White House complex joined together to suggest the use of evaluation is increasingly valuable during periods of constrained fiscal environments and to help agencies reach budget decisions (OMB, OSTP, DPC, & CEA, 2013).

The Obama Administration initially proposed a modest evaluation initiative that would (1) improve evaluation transparency, (2) create a *working group* to promote evaluation, (3) provide additional funds for agencies that voluntarily sought to improve production of impact evaluations (Brass, 2011; OMB, 2009b). These early efforts served as the underpinning of the evaluation capacity building strategy that would evolve to more broadly focus on (1) transparency of evaluation production and communication of results, (2) training and collaboration among agency staff, and (3) addressing barriers that limit the production of Federal

program evaluations. The Administration also pursued further the idea that funding should be and can be prioritized based on areas with existing evidence to guide decision-making. The Administration's emphasis on improving evaluation capacity started broadly, but over time focused more closely on social, education, and economic programs aimed at improving life outcomes.

Evaluation Transparency and Tiered Evidence. In response to criticisms that government sometimes conducted evaluations that were never made publicly available, the Obama Administration committed to publicly providing lists of evaluations in agencies to ensure results were publicly disseminated and to facilitate the eventual use of evaluation results. In the months following the first evaluation guidance, the Obama Administration announced additional government-wide direction to promote transparency, participation, and collaboration in an Open Government Directive (OMB, 2009d). The effort to increase awareness for new or ongoing evaluations, as well as existing evaluations, did lead to some apparent successes within a short period of time. For example, the Environmental Protection Agency launched a new website on program evaluation that at the time outlined both completed and ongoing evaluation activities (EPA, 2014), the Department of Labor launched the Clearinghouse for Labor Evaluation and Research (DOL, 2014b), and the Department of Justice's Crimesolutions.gov launched in June 2011 (NIJ, 2014). Other programs, such as HHS's Administration for Children and Families (ACF) took the opportunity to commit to evaluation policies that satisfied principles of transparency, independent, and policy relevance (HHS, 2014).

Efforts to improve the transparency and public dissemination of evaluations resulted in some work by OMB and staff in social agencies to develop a common evidence framework (OMB, 2013a). The work evolved from proposals to prioritize funding grantees that

implemented programs supported by some level of evidence, with additional weight placed on interventions supported by causal impact evaluations (OMB, 2012a; OMB et al., 2013). Such common evidence standards supported by OMB were drafted and made available in the clearinghouse websites supported by HHS (HHS, 2013), Labor (DOL, 2014a, 2014b), and Education (IES, 2014). Some non-profit organizations also attempted to further the use of these clearinghouses by creating techniques for both encouraging the increased development of clearinghouses (America Achieves, 2013) and integrating the various websites and rating schemes in to a single analytical tool (Pew, 2014).

The Administration's pursuit of a three-tiered evidence approach sought to incorporate evidence requirements in certain grant programs. The first tier is that initiatives supported by stronger evidence would receive greater amounts of funding. Second, programs with some, limited evidence received funding conditional on the programs improving evaluation efforts in the future. Third, agencies were encouraged to test ideas with potential when supported by research findings or "reasonable" hypotheses (OMB, 2010d). The example for this design is the Department of Education's Invest in Innovation Fund (i3). Applicants for i3 receive funding for a variety of purposes including developing, validating, and scaling programs. The amount of funding grantees receive depends on the evidence tier so as to create the "right incentives" for grantees in the future (OMB, 2010d). The tiered evidence approach eventually led OMB and several Federal agencies, notably Labor, Education, and the Corporation for National Community Service, to develop competitive grant designs that incorporated some element of evidence requirement in determine the allocation of funds. One early assessment of the tiered evidence approaches used for teen pregnancy prevention, home visiting, and other social innovations, written by advocates of the Obama Administration's approach, said that the

experimental strategy is promising but it is still too early to know whether the strategy was successful in deploying evaluation results to create a meaningful feedback loop for decision-making (Haskins & Margolis, 2014).

Agency Collaboration, Training, and Leadership. In 2009, OMB created an interagency working group within the continued PIC created by the Bush Administration to focus on strengthening agency capacity, sharing best practices, and provided technical assistance and research expertise to Federal agencies (OMB, 2009b). While some external analysts speculated the working group might formally direct evaluation practices (Brass, 2011), the implementation of the group more closely aligned with identifying areas where agencies could use additional technical assistance or networking among evaluators within the executive branch (OMB, 2012b). Operationally, the efforts appear to have been limited in scope and breadth of agency personnel involved. This group was eventually recast as a temporary series of forums and workshops for Federal agency staff to share best practices and begin to identify techniques to integrate performance measurement and evaluation (Keegan, 2014; OMB, 2013a, 2013c; OMB et al., 2013; Stack, 2014). The efforts by OMB to engage in training of agency staff were limited by OMB's own staff capacity and competing priorities for their time.

Public evaluation efforts also emphasized the need to strengthen evaluation capacity with high-level policy officials to support the program evaluation function. OMB offered a suite of policy options for agencies to consider from chief evaluation officers to empowering other agency officials (OMB, 2012c). Over the course of the Obama Administration, the Department of Labor, the Centers for Disease Control, and the Corporation for National Community Service all created new chief evaluation officer positions.

In 2015 OMB announced the creation of a new Leaders Delivery Network to focus on the implementation of agency priority goals. The effort is described as a strategy to encourage senior leaders across Federal agencies to share insights, expertise, and connections, providing access to "thought leaders" in the field (Donovan, 2015).

Addressing Barriers to Evaluation. The Obama team attempted to address a series of commonly cited barriers to conducting evaluations of public programs—or at least acknowledge certain challenges—specifically the availability of resources and flexibility for contracting, limitations on conducting surveys and information collection in the Federal government, and the availability of administrative datasets.

Increasing Evaluation Funding and Flexibility. In 2009, the Administration announced a \$100 million evaluation initiative to provide additional resources to 17 agencies to either support specific evaluations or increase agency evaluation capacity (OMB, 2009b, p. 5; 2010d). While some of these proposals may have been funded by Congress, no detail is readily available to indicate how many evaluations in the proposal were ultimately funded and undertaken by Federal agencies.

While early efforts focused on providing resources to fund specific evaluations, later strategies employed by the Administration focused more on increasing funding set-asides for existing programs. There were some notable successes in increasing funds though, such as the joint Social Security Administration (SSA) and Education project to evaluate interventions for improving outcomes for youth receiving Supplemental Security Insurance (SSI) benefits. The Administration has continued to request funding increases in other programs too. For example, the fiscal year 2016 budget requests authority from Congress to provide the Social Services Block Grant (SSBG) at HHS and the Department of Labor with the ability to allocate up to one

percent of program funds to support research and evaluation, in addition to investments in child care, child support, and criminal justice research (OMB, 2015).

The Obama Administration pursued other efforts to improve the flexibility of availability funds in government contracts. The Federal government's procurement processes can often be difficult to navigate, even for Federal employees, and may result in some amount of inefficiency for funds otherwise appropriated for evaluation purposes. For example, funds not allocated to projects within a given timeframe may become unavailable for assignment to other ongoing projects. Because evaluation activities tend to occur over multiple years, the absence of flexibility in the use of funds can be limiting for conducting well-planned evaluation activities. In 2015, the Obama Administration proposed for the first time to provide flexibility to Labor, Education, DOJ, HHS, Commerce, and HUD to pilot streamlining procurement processes for evaluation that extend the availability of funding to accommodate evaluation-specific contracting processes and timelines (HHS, 2015a; OMB, 2015).

Improving Survey Clearance Protocols. One potential impediment to Federal evaluation cited by practitioners is the limitation from the Paperwork Reduction Act (PRA) implemented by OMB to obtain a formal clearance for any survey with 10 or more participants. The process is intended to limit the burden imposed on the public in responding to government requests for data (OMB, 2009c), but evaluators have identified that the process can delay timeliness of evaluations and generate unintended consequences with those delays (Rog, Trochim, & Cooksy, 2009b). In 2009, OMB's Office of Information and Regulatory Affairs (OIRA) solicited public comments on ways to improve the implementation of PRA (74 FR 206, 55269). While the evaluation community viewed the opportunity to provide feedback on the process as a potential opening for

improvements, no formal action has since been announced by the Administration in response to the received public comments.

Providing Access to Administrative Data. A third barrier the Obama Administration sought to address in the evaluation domain was to provide researchers and evaluators greater access to administrative datasets collected by the Federal government. Early efforts successfully initiated access to administrative data through websites like data.gov and for use in program evaluations and priority goal activities (OMB, 2012b; U.S., 2014). Recognizing substantial barriers to these datasets persists, the fiscal year 2016 budget proposes to allow access to unemployment insurance data for measuring employment outcomes, requests additional funding to build technological infrastructure to improve access to state administrative datasets for Federal programs, and supports efforts to establish a commission that would recommend other ways to improve access to administrative data (OMB, 2015). A similar commission was first considered in Congress and is expected be re-proposed in 2015 (Clark, 2015).

Rapid Cycle Evaluation

Through the White House's Office of Science and Technology Policy, the Obama Administration created a Social and Behavioral Sciences Team (SBST) in 2014 to apply lessons from behavioral economics to Federal programs, modeling on a similar program in the United Kingdom. The initiative was intended to encourage Federal agencies to test the effects of informational and behavioral interventions in government programs on short-term outcomes. An annual report from the group in 2015 cites that the first year of the initiative yielded benefits specifically in streamlining program access and improving the efficiency of government programs through small scale projects at a variety of agencies (OSTP, 2015).

In September 2015 the Obama Administration started the process of institutionalizing the SBST efforts. President Obama signed an executive order directing Federal agencies to apply behavioral insights in a manner that better provides government services to the American public (Obama, 2015). In conjunction with the executive order, a consortium of non-governmental institutions and academic research centers committed to developing white papers by late 2016 that will propose specific uses of behavioral science research in the Federal government, with appropriate evaluation support (White House, 2015).

Encouraging Innovation

The Obama Administration also advocated for a number of initiatives that support evidence-based policy goals, even though the concepts diverge somewhat from evaluation of existing programs to support more innovative or forward-leaning activities.

Idea Labs and Innovation Funds. In 2010, HHS launched an Idea Lab, intended to provide Federal employees with innovative ideas with the resources to pilot test a demonstration project with appropriate evaluation of the strategy. The effort has continued since and generated proposals about connecting children to dental care, sequencing genomes of thousands of pathogens for public use, to devising technological improvements for tracking organ donations (HHS, 2015b). The Administration suggested expanding the initiative to five other Federal agencies (OMB, 2015). Various other initiatives supported the creation of innovation funds to try out new ideas in programs that improve implementation and service delivery, or reduce the costs to taxpayers. The Investing in Innovation Fund (i3) at the Department of Education received initial funding in 2009 in the American Recovery and Reinvestment Act (ARRA). That fund's purpose is to identify innovative practices for closing achievement gaps in education, decreasing dropout rates, and improving graduation rates (Education, 2015b). The Administration's support

for innovation funds appeared consistent with other proposals related to energy assistance, child support, and global poverty in the fiscal year 2016 budget request (OMB, 2015).

Pay for Success. Beginning in 2011, the Obama Administration began advocating the use of Social Impact Bonds financing and the Pay for Success contracting model as a way to drive innovation for social program service delivery in the private sector. Under the PFS model, governmental organization fund contracts based on local providers' ability to achieve specified performance targets, increasing payment when performance greatly exceeds targets and declining payment when performance goals are not achieved (Liebman, 2011, 2014). While the idea has been around for decades, the basic premise was advocated particularly during the Clinton NPR initiative as performance-based contracting and extensive academic discussion has ensued about the challenges and potential risks of the mechanisms (Galloway, 2013). Nonetheless, the Obama White House advanced a strategy that fully supported the technique to better drive innovation in the non-profit and private sectors (OMB, 2011d).

Innovation Exchange. During the second half of the Obama Administration, an initiative led by the Department of Education called the "Innovation Exchange" sought to spread ideas and examples throughout Federal agencies with semi-regular events small groups or teams from various agencies (Education, 2015a). While not coordinated by OMB, the Innovation Exchange was designed to spur new ideas through a decentralized learning forum. The forum described many of the strategies applied or attempted in other agencies that related to evaluation capacity and performance measurement improvements.

Beyond the many proposals and initiatives described here, an increasing number of other Administration initiatives have begun embedding data standards and evidence requirements into their designs, although in many cases it may still be too early to know what practical effect such

activities will have on evaluation capacity and performance management. Piecing together the evolution and effect of the Obama Administration's performance measurement and evaluation strategies to provide a narrative reveals a complex, multi-faceted agenda for improving government performance.

V. Lessons Learned and Re-Learned in the Obama Administration

In contrast to the various reviews of the Bush Administration's initiatives, which now have the benefit of hindsight, few in-depth assessments have been produced to date on the Obama Administration initiatives. The following section considers the Obama initiatives in context based on progress to date, and reflects on what lessons were learned by staff in the executive branch relative to insights gained from the initiatives launched during the Bush Administration.

- **Purpose.** The Obama priority goal effort was slow to get off the ground and evolved substantially over time to reflect changing political and administrative conditions. Two core challenges faced the priority goal effort from the outset: articulating the broader agenda within which they would be situated and what role they would play within that agenda. Separately, the evaluation reforms fell into an amorphous evidence-based policy agenda, yet the precise purposes for the reforms oscillated between serving a role for budget cuts and at other times for program learning and improving program effectiveness. A cohesive framework for the various evaluation initiatives never appeared to materialize, and emphasis was minimal in programs outside social policy. The Bush Administration faced similar challenges with implementing the PMA, where the reforms aimed to fill many related, yet distinct purposes, perhaps too many to be the most clearly and effectively implemented.

- **Accountability and Transparency.** The Obama Administration seized the opportunity to promote transparency in both performance measurement and evaluation initiatives throughout the administration. However, the ability to translate the transparency efforts to public or congressional accountability is less clear. The Obama initiatives prioritized providing information about performance and results, sometimes to the point inundation, but information users needed to have an ability to meaningfully filter or translate the information. Building on the efforts of the Bush Administration to provide performance information and some evaluations on the internet, the Obama Administration continued moving such efforts in that direction. Architects of the Obama policy previously acknowledged that different audiences have different information needs (Metzenbaum, 2009). By focusing priority goals on the public and relegating other performance information to the Annual Performance Plans and long documents, this idea may have been reasonably implemented. The Obama OMB had already committed to building a website highlighting performance data similar to Canada's portal before Modernization Act required it, but performance.gov has been criticized for insufficiently addressing intended audiences (GAO, 2013).
- **External Stakeholder Support.** At the beginning of the Obama Administration congressional dynamics tended to favor White House policies since the Democratic Party controlled both Congress and the White House. In that sense, passage of GPRA Modernization and support for some early evaluation initiatives signaled a level of interest in performance and evaluation not seen during the Bush Administration. Yet, GPRA Modernization did not require outreach to congressional committees in identifying and setting priority goals. Unlike the dynamic in the Bush years where budgetary

consequences were minimal in the absence of congressional interest, GPRA

Modernization provided certain additional reporting requirements for agencies if goal targets were not achieved. This lesson materialized less as a way of ensuring explicit support from external stakeholders than an accountability measure for senior officials.

- **Senior Official Engagement.** Early efforts to establish priority goals and GPRA Modernization elevated the role of senior executives who were responsible for achieving priority goal targets as well as senior policy officials in agencies for monitoring performance for priorities. In evaluation initiatives, in select agencies new chief evaluation officer positions empowered some individuals to support more and stronger evaluation policies. But senior officials were never required to utilize or pursue evaluation efforts, and many in areas outside the social policy domain never did. That said, senior White House officials were engaged early in the Obama administration in supporting the so-called evidence agenda and helped push through some of the necessary statutory changes and funding bills with provisions favorable to the initiatives, including the Pay for Success model and certain innovation funds (Haskins & Margolis, 2014). The high level engagement by senior White House officials likely contributed to the initiative's endurance over the years.
- **Workload Expectations.** Priority goals shifted considerable workload away from OMB program examiners and staff who under the Bush Administration had been responsible for both helping to establish metrics, set targets, and engage with agencies in setting stretch goals. Under the priority goal scheme, agencies led the effort to identify goals and to set targets for the majority of the goals. At the same time, the burden on agency staff was substantially lessened from the PART process where senior executives and staff

expended considerable resources preparing and defending PART self-assessments to OMB staff. All-in-all the collective workload from Obama performance and evaluation efforts appears to be lessened.

- **Uniformity-Flexibility Trade-off.** In addressing the criticisms of the PART's uniformity, the Obama approach allowed agencies considerable flexibility in defining their priority goals. Some might suggest the degree of flexibility was too far, though, with agencies unwilling to establish stretch targets that would have been advocated by OMB staff under the Bush Administration approach. Within the evaluation initiatives, agencies were also given considerable flexibility to voluntarily participate in virtually every aspect of the initiative, whereas PART equally challenged a wide range of programs to participate in evaluation efforts. Stepping back from the PART's commitment to make evaluation ubiquitous, the Obama evaluation initiatives were targeted to social programs in a more limiting manner, even though the marginal gains from establishing evaluation cultures in other agencies and policy domains could have been substantial.
- **Desired Goal Level.** The Obama approach to considering the level of performance activities elevated goals from the granular program level employed in the Bush strategy. The Obama initiatives returned the goal setting focus to the agency level, and envisioned that individual political appointees would participate in the goal setting exercises to align interests with the President's priorities. In contrast, PART efforts had worked to establish thousands of metrics that rolled in to agency strategic goals in a sometimes incoherent manner. While fewer goals were offered in the Obama Administration, those that were focused more on short-term achievements rather than long-term ones. This strategy may

limit benefits from pursuing longer term policy goals that may not neatly align with campaign cycles. Generally the Obama initiatives relied on setting strategic goals and objectives, which aligned with agency strategic plans, in addition to priority goals.

- **Staff Training.** OMB did engage agencies in training on evaluation, though the efforts were inconsistent and never intended for a broad audience of federal evaluation staff. On performance measurement issues, no evidence of a substantial training is available for either OMB staff or agency staff. However, the Innovation Exchange ideas advocated by agencies reportedly encouraged training and dissemination of capacity practices across agencies.
- **Linking Performance and Evaluation.** Administration guidance, OMB staff, and many agency operations continue to bifurcate the issues of performance measurement and program evaluation. While the Bush Administration acknowledged some role in incorporating the dynamics within the PART instrument, little effort appears to have been made in the Obama Administration to reconcile the integration of the two fields. To some extent, the practices may be further apart today than at the beginning of the Obama term due to the creation of a separate stand-alone operation on evaluation within OMB's career staff. During the Obama Administration, OMB created the informal "Evidence and Innovation Team" within its Economic Policy office to coordinate evaluation capacity building activities, while the Performance Management office separately designed and coordinated the management and priority goal activities. And the SBST staff was established in yet another component within the Executive Office of the President (EOP), the Office of Science, Technology, and Policy (OSTP). The separation of activities

within the EOP only mirrors the structure of many Federal agencies which similarly maintain distinct operations to coordinate evaluation and management functions.

- **Defining Quality of Evidence.** Like the Bush initiatives, the Obama efforts continued to apply a definition of evidence quality that aligns with positivist views that high-quality, rigorous evaluation is synonymous with RCTs. The perspective was prevalent for several years until a point where Obama administration guidance subtly referenced qualitative studies and other types of evaluation as valuable. Yet, RCTs were nonetheless prioritized in virtually all discussions about evaluation. More recent efforts to establish tiered frameworks for the quality of evidence also continue to prioritize causal randomized studies over alternative techniques that may provide complementary evidence of external validity or construct validity that is not provided by impact evaluations in isolation.
- **OMB's Role.** Obama's performance efforts were structured very differently from the Bush administration approaches which placed OMB staff in an authoritative role with the burden of proof on agencies for justifying their actions. Under the Obama approach, OMB's role was to design and globally coordinate the initiative, without interjecting at a micro level to apply a heavy hand in setting goals or targets. While OMB is more involved in coordinating and reviewing cross-agency goals, the emphasis on implementation and attainment continues to rest with the agency staff. Similarly, for evaluation initiatives OMB's role in the Obama era was largely facilitative, providing resources to agencies willing to participate in various efforts. One important feature of the OMB efforts in both administrations is that the processes were incorporated to some degree within existing budget and policy review processes. While in recent years some of the budget processes have become irregular, the general consistency of such processes

provides a defined infrastructure for decision-making that can connect initiatives with much needed action forcing events for both performance and evaluation activities.

The Obama approaches to implementing priority goals and the evaluation initiatives benefited in some cases from the Bush Administration's experiences with similar efforts. In other cases, the Obama Administration changed course based on observations from the Bush Administration's lessons. In still other areas, similar challenges and constraints persisted.

VI. Overarching Observations and Lessons

Attempts to reform Federal performance measurement and evaluation systems under the Bush and Obama Administrations have substantial similarities in fundamental approaches, and both strategies have yielded modest results, for a number of similar reasons. In fact, while it seemed at first that the Obama approach had benefited from lessons learned during the Bush Administration, in fact, many of the same challenges appear to have continued. We identify eight remaining challenges that neither administration fully addressed that relate to the lessons noted above for each administration.

1. Calibrating OMB's Role vis-à-vis the Agencies

All of the initiatives previously described struggled to appropriately calibrate the balance in the role of OMB vis-à-vis the executive agencies regarding ownership of the performance or evaluation efforts. OMB has an institutional responsibility to implement the president's agenda, and therefore maintains the power and influence to ensure agency staff are attentive to prescribed requirements. For example, OMB can enforce requirements on agencies to provide PART self assessments and performance.gov data. But when agency perspectives suggest OMB staff will render judgment on the quality or adequacy of analyses—particularly when agencies perceive the OMB judgment criteria as narrow or inconsistent with agency missions and needs—these

processes can and often do become merely compliance exercises where agencies minimally participate. Take, for example, OMB's emphasis on impact evaluation within agencies that do not have basic evaluation capacity or familiarity with precursors such as logic models and process evaluations. In such cases, agencies will almost always have difficulty "owning" new initiatives.

2. Establishing and Sustaining an Audience

Establishing and then maintaining the attention of an implementing audience for the performance or evaluation initiatives has been challenging for both the Bush and Obama administrations. In both instances, Administration initiatives became linked closely with certain political appointees within OMB, then when those appointees depart the initiatives become less urgent for agencies. This may, in part, be due to shifting priorities within OMB, but the effect on agencies is clear. Further, OMB's small staff generally lack the capacity due to competing demands to ensure initiatives receive constant attention and appropriate outreach and technical assistance for agency staff charged with translating OMB's guidance. While OMB has engaged in trainings and workshops through the PIC and several other forums, such efforts reach small audiences and tend to focus on only evaluation or only performance.

3. Effectively Implementing Initiatives with Multiple Agencies

Implementing government-wide, or cross-agency initiatives effectively to improve government performance has been a repeated challenge. Cross-agency collaboration is sorely needed, yet it has long been an oxymoron. To ensure that effective and authentic collaboration happens, substantial funding and staff time are required. The Obama Administration took steps in this direction in its FY 2016 Budget, requesting a small amount of funding to help coordinate

cross-agency priority goals (OMB, 2015), but agencies will need to fully collaborate on complex issues that span intergovernmental jurisdictions to be the most effective.

4. Generating Sufficient Success Stories.

Neither administration developed and disseminated sufficient success stories to relay to Federal executive leaders, Congress, or the public that the performance or evaluation initiatives “worked.” Because leadership for performance or evaluation initiatives requires persuasion, sharing of success stories can be beneficial. But administration representatives rely on the same handful of stories, which may be generous representations of reality, and agency audiences and even OMB staff can easily become skeptical of the value of the initiatives.

5. Building Adequate Evaluation Capacity.

Building adequate evaluation capacity in agencies beyond Education, Labor, and HHS is necessary to support performance and evaluation initiatives over the long-term. Basic capacity can help agencies identify what data to collect and to appropriately analyze the data to learn important lessons about program implementation and improvement. Both administrations predominately focused on RCTs for the purposes of impact evaluation. Such an approach alienated staff in agencies beyond the handful of agencies with sufficient resources and expertise that historically administered RCTs. Reliance on and calls to increase the use of RCTs have been at the expense of mixed methods for evaluations with formative purposes.

6. Establishing Synergies Between Evaluation and Performance Staff.

Additional emphasis is needed on creating and institutionalizing synergies among performance measurement and evaluation offices and staff within the agencies. In virtually all Federal agencies, including Labor and HHS, performance reporting occurs in totally separate offices from evaluation initiatives. The two activities are viewed by agency staff as separate

enterprises, with little interaction, thus performance measurement and reporting do not benefit from nor feed into evaluation initiatives. A separation between performance measurement, on one hand, and the rest of evaluation practice, on the other, has also been manifest for years in OMB and GAO, both organizationally and in their written products. Addressing this challenge will be difficult because bureaucratic inertia, cultural norms, and individual worldviews are difficult to modify.

7. Socializing and Training Political Appointees and Senior Executives Appropriately.

Neither administration made notable strides in socializing and training political appointees to assume visible and authentic leadership roles in performance or evaluation initiatives. With typically short tenures, political appointees barely have time to learn the ins-and-outs of the agencies' and programs' missions and then try to “make a difference.” Performance or evaluation initiatives are typically not among new political appointees' highest priorities. Targeting training at political appointees through the White House's Office of Presidential Personnel could help address this challenge. In addition, greater exposure to the initiatives among career senior executives who work and consult directly with senior political officials could also contribute to addressing the challenge. OMB has done little to create robust evaluation skillsets among senior career executives, which in turn means that political appointees may often not receive sufficient support to implement the initiatives.

8. Consulting with Appropriate Congressional Leaders and Staff.

Neither the Bush or Obama Administration consulted with congressional leaders about the design of performance and evaluation initiatives and in turn congressional staff did not appear to make use of information made publicly available for Congress. Consulting with interested congressional committees or leaders about the need for or potential value of proposed

performance or evaluation initiatives during the design phase would help ensure the initiatives are appropriately calibrated to meet congressional needs. Presidential management agendas or frameworks, including performance or evaluation initiatives, are just that – they are owned by the President and the executive branch. While GPRA and the GPRA Modernization Act specifically called for more congressional consultation, to date there has not been a strong track record of such collaboration occurring even when there are not ideological differences among the branches. The breakdown in consultation is perhaps more pronounced when different political parties control the two branches.

In the era of so-called evidence-based policymaking, the focus on incorporating additional information into decision-making processes is closely linked to the idea prevalent over past decades that policymakers are able to and capable of identifying such information and the appropriateness for applying such information in a decision-making framework. At the same time the performance measurement and evaluation efforts discussed above have sought to both separately provide some level of accountability, describing the current effectiveness and performance of government programs. But determining what constitutes valid, relevant, and reliable evidence for many Federal programs and agencies is complicated for those staff operating the programs—typically more challenging than anticipated at the outset—and can be even more difficult to interpret for individuals with expansive oversight functions, such as congressional committees and OMB staff.

Capacity to improve the production of performance measurement and evaluation activities remains limited if not inconsistent across the Federal government. The U.S. government could continue to explore ways to better develop the capacity to align with the dual purposes defined for pursuing these activities. Efforts to develop baseline assessments of

capacity to determine agency starting points for these activities may offer a clear method for OMB and other coordinating offices to better target assistance and resources for improving overall capacity in support of identified evidence-based priorities. Similarly, the use of PART to require evaluation as a mandated question for programs encourage the production of new evaluations but may have relayed evaluation requirements in an overly uniform manner that did not accommodate different evaluation methods other than RCTs, which some programs may not have been prepared to or capable of implementing.

Future attention could also help operationalize the theoretical interrelationship between performance measurement and evaluation activities. In monitoring and evaluation accountability systems, indicators of program performance can highlight projects meeting defined targets, provide indicators for impact assessments, and fulfill overarching accountability requirements (Patton, 2008). While not suitable for all programs, an acknowledgment sometimes missing in recent administrations, performance measures can provide quality information for program managers about the direction of outputs and productivity, as well as trends over time, with a periodicity necessary for managers and policymakers to make course corrections in a timely fashion (McDavid & Hawthorn, 2006). However, while performance monitoring can provide critical signals about intervention efficacy, performance measurement is often criticized for a lack of ability to explain why trends are occurring or to attribute causal effects to specific programs (Hatry, 2013). Process evaluations, among other methods, can provide insights about performance monitoring information for how interventions operate and why programs are operating in the specified manner. Then, other evaluation techniques can then be employed to connect indicators to interventions and explain how effects interact or impact the desired program outcomes.

Rarely is performance monitoring and measurement envisioned as an activity that occurs in isolation from program management and decision-making. Performance monitoring is frequently described as just one of a series of activities to improve overall program performance. Wholey (2012) notes that monitoring and evaluation positively reinforce each other and that evaluations can become more useful and less costly when programs have repeatedly collected performance data and refined performance metrics to improve their performance.

In practice, disconnects between the two functions have persisted across administrations and through countless initiatives. The U.S. Federal government's implementation of performance measurement and evaluation within a cohesive performance management framework is tenuous as the predominance of performance measurement activities maintains a foothold in Federal agencies. Current practice within the Federal government tends to emphasize performance measurement systems where the two fields often operate in silos and where practitioners of evaluation and performance measurement are assigned to different organizational units. Following enactment of GPRA, evaluators were rarely relied on to inform performance measurement decisions (Newcomer, 1997). Similar observations have been made during the Bush and Obama Administrations.

The Bush Administration sought to apply the PART to provide an overarching performance and analytical framework, although PART featured far greater emphasis on output and efficiency performance measures than evaluation. The Obama Administration, which replaced PART with an alternative framework, similarly espoused the benefits of integrating the complementary activities all while adopting a management framework that largely continued the bifurcation of performance measurement and evaluation activities (OMB, 2009a, 2010b, 2010d, 2011e). The predominance of performance measurement in the Federal government, and its

periodic misuse in decision making processes, may contribute to the continued divergence in the activities as evaluators seek to maintain some distance from performance monitors, and vice versa (Nielsen & Hunter, 2013). Further, the prevalence of performance measurement activities may also be directly related to the immediate information needs of decisionmakers, who may not have sufficient timeframes to wait for retrospective insights from evaluations that take long periods of time to produce. While the theoretical importance for pursuing integration is evident, the practical challenges to integrating monitoring and evaluation activities are numerous.

Even when one function is described as critical for the other, the day-to-day operations within agencies and OMB result in strategies that fall short of integration. One path forward might be to encourage agencies to better integrate the functional positions and offices within agencies that oversee evaluation and performance efforts, beginning with OMB. In the mid-1990s, OMB recognized the importance of management functions in coming years and reorganized the entire agency to integrate certain management functions within the budget and policy operations (OMB, 1994). While integration could result in competing priorities crowding out the less established function, coordination of the two efforts seems unlikely to move beyond distant complementarity without actions for the functions to work in tandem at the coordinating and operational levels (Newcomer & Brass, 2015).

Finally, implementation lessons are abundant from the efforts over the past two administrations. Administrations quickly learn it is one thing to announce new initiatives and another to implement them, and to do so well. While some lessons are important irrespective of strategy, the value of others rests with the vision and strategy for use outlined by the respective administration as well as identifying an appropriate balance between uniformity, flexibility, centralization, and decentralization. For instance, training, infrastructure, and sufficient staffing

are likely essential elements to successful implementation of any initiative that seeks to introduce new concepts or operations to agencies. But above all the value of engagement of senior leadership within agencies and among external stakeholders is paramount to fulfilling any “evidence-based” initiative that seeks to incorporate evidence in to a decision-making process. For engagement to occur, the implicit or explicit theory of evaluation use is crucial. If stakeholders are presented with evaluation methods or information they view as biased or too narrow to be useful, they will not use the information. Without buy-in from senior leaders who perceive information as useful and who are willing to utilize the information developed through performance measurement and evaluation efforts – the very definition of evidence-based policymaking -- the initiatives success will almost always be futile. Other efforts to integrate collected performance data in to decision frameworks, with appropriate context, should be encouraged.

While the Obama Administration's performance measurement and evaluation capacity building strategies continue to evolve, the Bush Administration's efforts provided lessons and some structure for initiatives the continued or launched in the Obama term. This paper has highlighted many similarities across the administrations, a number of differences, and focused on lessons learned from different aspects of the experiences, offering a series of persistent challenges to be addressed in the design of future administrations' initiatives.

References

- America Achieves. (2013). Investing in What Works Index. Retrieved May 9, 2014, from <http://www.americaachieves.org/docs/InvestInWhatWorksIndex.September2013.pdf>
- Barzelay, M. (2001). *The new public management: improving research and policy dialogue* (Vol. 3). New York: University of California Press.
- Bernholz, E., Carra, J., Clark, P., Ginsburg, A., Habibion, M., Heffelfinger, J., . . . Valdez, B. (2006). *Evaluation Dialogue Between OMB Staff and Federal Evaluators* Washington, D.C. Retrieved from http://georgewbush-whitehouse.archives.gov/omb/performance/guidance/part_appeals_guid_2008-03.pdf
- Brass, C. T. (2011). *Obama Administration Agenda for Government Performance: Evolution and Related Issues for Congress*. Washington, D.C.: Congressional Research Service Retrieved from <http://fas.org/sgp/crs/misc/govperf.pdf>.
- Brass, C. T. (2012). *Changes to the Government Performance and Results Act (GPRA): Overview of the New Framework of Products and Processes*. Washington, D.C.: Congressional Research Service.
- Breul, J. D. (2007). Three Bush Administration Management Reform Initiatives: The President's Management Agenda, Freedom to Manage Legislative Proposals, and the Program Assessment Rating Tool. *Public Administration Review*, 67(1), 21-26.
- Breul, J. D., & Kamensky, J. M. (2008). Federal Government Reform: Lessons from Clinton's "Reinventing Government" and Bush's "Management Agenda" Initiatives. *Public Administration Review*, 68(6), 1009-1026.
- Clark, C. S. (2015). Bipartisan Duo to Reintroduce Bill on Evidence-Based Policymaking, *Government Executive*. Retrieved from

<http://www.govexec.com/management/2015/02/bipartisan-duo-reintroduce-bill-evidence-based-policymaking/105393/>

Consolidated Appropriations Act, 2008, Public Law 110–161 § Section 519 (2008 December 26, 2007).

Dahler-Larsen, P. (2011). *The Evaluation Society*: Stanford University Press.

DOL. (2014a). *CLEAR Causal Evidence Guidelines, Version 1.1*. Washington, D.C.: Retrieved from http://clear.dol.gov/sites/default/files/CLEAR_EvidenceGuidelines_1.1_revised.pdf

DOL. (2014b). *Clearinghouse for Labor Evaluation and Research (CLEAR)*. Washington, D.C.: Retrieved from <http://clear.dol.gov/>.

Donovan, S. (2015). *Continuing to Drive Mission Results for the American People*. Washington, D.C.: White House Office of Management and Budget (OMB) Retrieved from <https://www.whitehouse.gov/blog/2015/10/08/continuing-drive-mission-results-american-people>.

Education. (2015a). *Innovation Exchange 2015*. Washington, D.C.: Retrieved from <http://www2.ed.gov/about/offices/list/ods/evidence/index.html>.

Education. (2015b). INVESTING IN INNOVATION FUND (I3) from <http://www2.ed.gov/programs/innovation/>

EPA. (2014). EPA Program Evaluation Reports. Retrieved May 10, 2014, from <http://www.epa.gov/evaluate/reports/index.htm>

Galloway, I. (Ed.). (2013). *Pay For Success Financing* (1 ed. Vol. 9).

GAO. (2004). *PERFORMANCE BUDGETING: Observations on the Use of OMB's Program Assessment Rating Tool for the Fiscal Year 2004 Budget*. (GAO-04-174). Washington, D.C.: GAO Retrieved from <http://www.gao.gov/assets/250/241315.pdf>.

- GAO. (2006). *PRESIDENTS MANAGEMENT AGENDA: Review of OMB's Improved Financial Performance Scorecard Process*. (GAO-07-95). Washington, D.C.: GAO Retrieved from <http://www.gao.gov/new.items/d0795.pdf>.
- GAO. (2013). *MANAGING FOR RESULTS: Agencies Should More Fully Develop Priority Goals under the GPRA Modernization Act*. Washington, D.C.: GAO Retrieved from <http://www.gao.gov/products/GAO-13-174>.
- Gore, A. (1997). *Businesslike government: lessons learned from America's best companies*. Washington, D.C.: National Performance Review.
- Gore, A. (1998). *Common Sense Government: Works Better and Costs Less: National Performance Review (3rd Report)*: DIANE Publishing Company.
- Haskins, R., & Margolis, G. (2014). *Show Me the Evidence: Obama's Fight for Rigor and Results in Social Policy*: Brookings Institution Press.
- Hatry, H. P. (2013). Sorting the Relationships Among Performance Measurement, Program Evaluation, and Performance Management. *New Directions for Evaluation*, 2013(137), 19-32.
- Heinrich, C. J. (2012). How Credible Is the Evidence, and Does It Matter? An Analysis of the Program Assessment Rating Tool. *Public Administration Review*, 72(1), 123-134.
- HHS. (2013). *Common Evidence Framework (Draft)*. Washington, D.C.: Retrieved from http://findyouthinfo.gov/docs/Common_Evidence_Framework-Draft_508_5-08-13.pdf.
- HHS. (2014). *Evaluation Policy*. Federal Register, Washington D.C.: GPO Retrieved from <https://www.federalregister.gov/articles/2014/08/29/2014-20616/evaluation-policy-cooperative-research-or-demonstration-projects>.

- HHS. (2015a). FY 2016 Budget Request: Administration for Children and Families. from https://www.acf.hhs.gov/sites/default/files/olab/2016_acf_cj.PDF
- HHS. (2015b). HHS Idea Lab. from <http://www.hhs.gov/idealab/>
- IES. (2014). What Works Clearinghouse. Retrieved May 9, 2014, from <http://ies.ed.gov/ncee/wwc/>
- Issenberg, S. (2012a, Oct. 29). Obama Does it Better, *Slate*. Retrieved from http://www.slate.com/articles/news_and_politics/victory_lab/2012/10/obama_s_secret_weapon_democrats_have_a_massive_advantage_in_targeting_and.single.html
- Issenberg, S. (2012b). *The Victory Lab: The Secret Science of Winning Campaigns*: Crown.
- Joyce, P. G. (2011). The Obama Administration and PBB: Building on the Legacy of Federal Performance-Informed Budgeting? *Public Administration Review*, 71(3), 356-367.
- Keegan, M. (2014). Harnessing Evidence and Evaluation: Insights from Kathy Stack, Advisor, Evidence-Based Innovation, Office of Management and Budget. *The Business of Government, Spring*, 48-51.
- Kettl, D. F. (1998). *Reinventing Government: A Fifth-year Report Card*. Washington, D.C.: The Brookings Institution.
- Liebman, J. (2011). Testing Pay-for-Success Bonds. *Public Manager*, 40(3), 66-68.
- Liebman, J. (2014). *Testimony of Jeffrey B. Liebman*. U.S. Senate Budget Committee Government Performance Task Force. Washington, D.C. Retrieved from http://www.budget.senate.gov/democratic/public/_cache/files/c9f5bc47-adf3-4b2b-bd9b-b8eff141c814/liebman-pay-for-success-testimony.pdf
- Mark, K., & Pfeiffer, J. R. (2011). Monitoring and Evaluation in the United States Government: An Overview: ECD Working Paper Series.

- McDavid, J. C., & Hawthorn, L. R. L. (2006). *Program Evaluation & Performance Measurement: An Introduction to Practice*: SAGE Publications.
- Metzenbaum, S. (2009). Performance Management Recommendations for the New Administration.
- Moe, R. C. (1987). Exploring the Limits of Privatization. *Public Administration Review*.
- Moynihan, D. (2010). *The politics measurement makes: Performance management in the Obama era*. Paper presented at the The Forum.
- Moynihan, D. (2013). Advancing the Empirical Study of Performance Management: What We Learned From the Program Assessment Rating Tool. *The American Review of Public Administration*, 43(5), 499-517.
- Moynihan, D., & Lavertu, S. (2012). Does involvement in performance management routines encourage performance information use? Evaluating GPRA and PART. *Public Administration Review*, 72(4), 592-602.
- Newcomer, K. (1997). Using Performance Measurement to Improve Programs. *New Directions for Evaluation*, 75(Fall), 5-14.
- Newcomer, K., & Brass, C. T. (2015). Forging a Strategic and Comprehensive Approach to Evaluation within Public and Nonprofit Organizations: Integrating Measurement and Analytics within Evaluation. *American Journal of Evaluation*.
- Newcomer, K., & Redburn, F. S. (2008). Achieving Real Improvement in Federal Policy and Program Outcomes *Obama-Biden Transition Project*. Washington, D.C.: National Academy of Public Administration and The George Washington University.

Nielsen, S. B., & Hunter, D. E. (2013). Challenges to and forms of complementarity between performance management and evaluation. *New Directions for Evaluation*, 2013(137), 115-123.

NIJ. (2014). CrimeSolutions.gov. Retrieved May 9, 2014, from <https://www.crimesolutions.gov>

Obama, B. (2009). *Remarks by the President on Community Solutions Agenda*. Washington, D.C.: The White House.

Obama, B. (2015). *Executive Order -- Using Behavioral Science Insights to Better Serve the American People*. Washington, D.C.: The White House Retrieved from <https://www.whitehouse.gov/the-press-office/2015/09/15/executive-order-using-behavioral-science-insights-better-serve-american>.

OMB. (1994). *Making OMB More Effective in Serving the Presidency: Changes in OMB as a Result of the OMB 2000 Review (M-94-16)*. Washington, D.C.: Retrieved from <http://www.clintonlibrary.gov/assets/DigitalLibrary/AdminHistories/Box%20041-050/Box%20046/1504319-omb-2000.pdf>.

OMB. (2001). *The President's Management Agenda*. Washington, D.C.: Retrieved from <http://www.whitehouse.gov/sites/default/files/omb/budget/fy2002/mgmt.pdf>.

OMB. (2003). *Budget of the United States Government, Fiscal Year 2004: Performance and Management Assessments*. Washinton, D.C.: Government Printing Office.

OMB. (2004). *What Constitutes Strong Evidence of a Program's Effectiveness*. Washington, D.C.

OMB. (2007). *Implementing Executive Order 13450: Improving Government Program Performance (M-08-06)*. Washington, D.C.

- OMB. (2008). *Guidance on Appeals Process for 2008 PARTs*. Washington, D.C.: Retrieved from http://georgewbush-whitehouse.archives.gov/omb/performance/guidance/part_appeals_guid_2008-03.pdf.
- OMB. (2009a). *Building a High-Performing Government Analytical Perspectives: Budget of the United States Government for Fiscal Year 2010*. Washington, D.C.: Government Printing Office.
- OMB. (2009b). *Increased Emphasis on Program Evaluations (M-10-01)*. Washington, D.C.
- OMB. (2009c). *Information Collection under the Paperwork Reduction Act* Washington, D.C.: Retrieved from http://www.whitehouse.gov/sites/default/files/omb/assets/inforeg/PRAPrimer_04072010.pdf.
- OMB. (2009d). *Open Government Directive (M-10-06)*. Washington, D.C.
- OMB. (2009e). *Planning for the President's Fiscal Year 2011 Budget and Performance Plans (M-09-20)*. Washington, D.C.: Retrieved from <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2014/m-14-07.pdf>.
- OMB. (2010a). *The Accountable Government Initiative—An Update on Our Performance Management Agenda*. Washington, D.C.: Retrieved from http://www.whitehouse.gov/sites/default/files/omb/memoranda/2010/AccountableGovernmentInitiative_09142010.pdf.
- OMB. (2010b). *Delivering High-Performance Government Performance and Management Analytical Perspectives: Budget of the United States Government for Fiscal Year 2011*. Washington, D.C.: Government Printing Office.
- OMB. (2010c). *Performance Improvement Guidance: Management Responsibilities and Government Performance Results Act Documents (M-10-24)*. Washington, D.C.

- OMB. (2010d). *Program Evaluation Analytical Perspectives: Budget of the United States Government for Fiscal Year 2011*. Washington, D.C.: Government Printing Office.
- OMB. (2011a). *Delivering an Efficient, Effective, and Accountable Government (M-11-31)*. Washington, D.C.
- OMB. (2011b). *Delivering on the Accountable Government Initiative and Implementing the GPR Modernization Act of 2010 (M-11-17)*. Washington, D.C.
- OMB. (2011c). *High Priority Performance Goals*. Washington, D.C.: Retrieved from <http://www.whitehouse.gov/sites/default/files/omb/performance/high-priority-performance-goals.pdf>.
- OMB. (2011d). *Paying for Success*. Washington, D.C.: Retrieved from <http://www.whitehouse.gov/omb/factsheet/paying-for-success>.
- OMB. (2011e). *Program Evaluation Analytical Perspectives: Budget of the United States Government for Fiscal Year 2012*. Washington, D.C.: Government Printing Office.
- OMB. (2012a). *Fiscal Year 2014 Budget Guidance (M-12-13)*. Washington, D.C.
- OMB. (2012b). *Program Evaluation and Data Analytics Analytical Perspectives: Budget of the United States Government for Fiscal Year 2013*. Washington, D.C.: Government Printing Office.
- OMB. (2012c). *Use of Evidence and Evaluation in the 2014 Budget (M-12-14)*. Washington, D.C.: Retrieved from <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2012/m-12-14.pdf>
- OMB. (2013a). *Budget of the United States Government, Fiscal Year 2014*. Washington, D.C.: Government Printing Office.
- OMB. (2013b). *Fiscal Year 2015 Budget Guidance (M-13-14)*. Washington, D.C.

OMB. (2013c). *M-13-14: Fiscal Year 2015 Budget Guidance*.

OMB. (2014). *Circular No. A-11: Preparation, Submission and Execution of the Budget*.

Washington, D.C.: White House Office of Management and Budget.

OMB. (2015). *Budget of the United States Government, Fiscal Year 2016*. Washinton, D.C.:

Government Printing Office.

OMB, OSTP, DPC, & CEA. (2013). *Next Steps in the Evidence and Innovation Agenda (M-13-*

17). Washington, D.C.: Retrieved from <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-17.pdf>.

Orszag, P. (2009). *Building Rigorous Evidence to Drive Policy*. Retrieved from

<http://www.whitehouse.gov/omb/blog/09/06/08/BuildingRigorousEvidencetoDrivePolicy>

OSTP. (2015). *Social and Behavioral Sciences Team Annual Report*. Washington, D.C.:

Retrieved from https://www.whitehouse.gov/sites/default/files/microsites/ostp/sbst_2015_annual_report_final_9_14_15.pdf.

Patton, M. Q. (2008). *Utilization-Focused Evaluation*: SAGE Publications.

Pew. (2014). *Results First Clearinghouse Database*. Retrieved 2014, from

<http://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2014/09/results-first-clearinghouse-database>

Rog, D., Trochim, W., & Cooksy, L. (2009a). *AEA Response to OMB re: Increased Emphasis on*

Program Evaluations American Evaluation Association. Washington, D.C. Retrieved

from <http://www.eval.org/d/do/98>

Rog, D., Trochim, W., & Cooksy, L. (2009b). *Improving Implementation of the Paperwork*

Reduction Act. American Evaluation Association. Washington, D.C. Retrieved from

<http://www.eval.org/d/do/104>

SAMHSA. (2014). National Registry of Evidence-Based Programs and Practices. Retrieved May 9, 2014, from <http://www.nrepp.samhsa.gov/>

(2013, March 27). *Amelia Showalter Interview on 'Learning from the Obama campaign about creating a culture of experimentation in government'* [Retrieved from <http://media.blubrry.com/govinnovator/p/govinnovator.com/wp-content/uploads/2013/03/Amelia-Showalter1.mp3>

(2014, March 30). *Kathy Stack Interview on 'What is Program Evaluation?'* [Retrieved from http://www.blochhousemedia.com/ibm/kathy_stack_abcd.mp3

Trochim, W. (2008). *EPTF Response to OMB re: What Constitutes Strong Evidence of a Program's Effectiveness*. American Evaluation Association Evaluation Policy Task Force. Washington, D.C. Retrieved from <http://www.eval.org/d/do/98>

U.S. (2008a). *ExpectMore.gov*. Washington, D.C.

U.S. (2008b). *Results.gov*. Washington, D.C.: Retrieved from <http://georgewbush-whitehouse.archives.gov/results/agenda/index.html>.

U.S. (2014). *Data.gov*. Washington, D.C.

U.S. (2015). *Performance.gov*. Washington, D.C.

White, J. (2012). Playing the Wrong PART: The Program Assessment Rating Tool and the Functions of the President's Budget. *Public Administration Review*, 72(1), 112-121.

White House. (2015). FACT SHEET: President Obama Signs Executive Order; White House Announces New Steps to Improve Federal Programs by Leveraging Research Insights [Press release]