Presidential Evidence Initiatives:

Lessons from the Bush and Obama Administrations' Efforts to Improve Government Performance

FEBRUARY 2018

BIPARTISAN POLICY CENTER

AUTHORS

Nick Hart, Ph.D.
Director of the Evidence-Based Policymaking Initiative
Bipartisan Policy Center

Kathryn Newcomer, Ph.D.
Director of the Trachtenberg School of Public Policy
and Public Administration
George Washington University

Executive Summary

Calls for assessing how well government programs work have been around for decades. 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 information.

This paper provides an overview of each administration's approach, then highlights four key lessons for improving frameworks for government performance and better enabling evidence-based policymaking in practice:

1. **Calibrate Evidence-Building Activities to Achieve Learning.** The calibration of evidence-building activities to meet common needs for program learning must be better articulated to support engagement.

2. **Coordinate Across the Evidence-Building Community.** Evidence initiatives must establish better connections across the evidence-building community to engage agency staff working on performance, evaluation, statistical, and other activities to leverage their strengths and generate synergies for producing relevant and timely evidence.

3. **Acquire and Sustain Audiences of Potential Users.** Government initiatives to improve program performance by using available evidence need to engage agency leaders and other evidence consumers to identify the types of information that will be most useful to inform decision-making.

4. **Implement Initiatives Effectively.** Evidence-building capacity of federal agencies should be strategically enhanced, and success stories about the use of evaluation and other relevant forms of evidence to improve agency performance should be systematically collected and shared.

The evidence initiatives over the past 20 years suggest that future initiatives aimed at improving government's effectiveness can certainly be better designed to integrate performance measurement and evaluation activities with the numerous other efforts within the evidence-building community. Doing so stands to improve how efforts to build evidence occur, and better inform the variety of uses among potential users.

Introduction

Since the 1990s, calls for increased government accountability, learning, and improved performance in the United States transformed how the federal government implemented performance management systems. These systems generally emphasize goal-setting, measurement, and reporting. More recent efforts to promote "evidence-based policy" have sought to steer resources based on the results of program evaluations. The first two U.S. presidents of the 21st century—George W. Bush and Barack Obama—both led administrations that encouraged these activities. Despite their different policy priorities, ideological perspectives, professional backgrounds, and senior advisors, both Bush and Obama established management and evidence agendas with common elements in their aim to promote government performance and public accountability.

Notwithstanding some differences between the Bush and Obama performance improvement and evaluation agendas, the two administrations' plans featured many similarities. The similarities may have contributed to strengthening overall implementation of reform efforts and paved a path for even more transformative efforts in the future. For example, both relied on a belief that by setting goals for government programs, improvements in outputs and outcomes are destined to follow; but that doesn't always happen in practice. Both also called for more widespread evaluation of program impact, though limited successes are evident today in practice. Both also failed to adequately bring together the siloes of individuals engaged in evidence-building across performance, evaluation, statistical, and research units. The approaches pursued by each administration highlight limitations that evolving approaches in public administration may be able to improve upon in the future.

Four overarching observations about the Bush and Obama performance management and evaluation agendas emerge for future efforts to consider: (1) many initiatives seek to serve both accountability and learning purposes, yet attainment of both goals is difficult, resulting in mandates that produce compliance-focused activities with a façade of meaningful use; (2) coordination across the entire evidence-building community can support generation of more relevant evidence and increase the likelihood of use among decision-makers; (3) acquiring and sustaining audiences to produce and use evaluation must strike the balance between supply and demand for evidence-building, though the circularity of various approaches has not identified which should come first; and (4) effective implementation of evidence-building capabilities has been constrained by institutions.

This technical paper outlines what evidence-based policy was intended to represent for the Bush and Obama administrations, then highlights the initiatives each pursued. The brief concludes by offering important lessons learned over the past two administrations about effectively implementing performance improvement initiatives.

"Results-Driven" to "Evidence-Based" Government

Following the inception of new major social investments during the mid-20th century, the field of public administration shifted increasingly toward market-oriented management practices and saw the advent of the New Public Management approach. By the 1990s, government reform advocates espoused principles for improving public programs based on case studies of government innovations, many of which served as the core principles for the Clinton administration's National Performance Review.[1] The Clinton philosophy was to take inefficient government programs and transform them to be more effective, more efficient, and more dynamic operations.[2] Thus, New Public Management ultimately helped increase the demand for more evaluation and production of data.[3]

By the late-1990s, implementation of the Government Performance and Results Act (GPRA) began to change the way federal programs reported on program performance, shifting from an emphasis on reporting program inputs to reporting information on outputs and, when possible, outcomes. GPRA directed agencies to improve planning processes through development of new strategic plans and annual performance plans. GPRA's advocates assumed that tracking performance would help agency leaders, Congress, and the public hold government accountable for achieving program objectives, and for continuously improving programs through analyzing performance information.[4] Shortly after enactment of GPRA, Congress reauthorized the Paperwork Reduction Act with a series of modifications that largely recognized goals of improving government efficiency, specifically in regard to the manner in which government imposes burdens on the public to collect information relevant for performance monitoring, evaluation, and other evidence-building activities.[5]

By some accounts the Clinton efforts succeeded in achieving discrete goals,[6] but by other accounts, they did not fully engage political leadership, improve trust in government programs, integrate the initiative with performance measurement, or consult with congressional stakeholders.[7] This perspective evolved in subsequent years as the Bush and Obama evidence agendas took hold. The emphasis became less focused on "results driven" and increasingly focused on "evidence-based" strategies, encouraging rigorous research and evaluation to be developed to supplement performance monitoring information.

Bush Administration Performance and Evaluation Reforms

Building on the efforts implemented during the Clinton Administration, when Bush came to office he 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.[8] To coordinate and oversee the PMA President Bush re-established the President's Management Council, which included agency chief operating officers. The coordinating body was intended to help ensure the sweeping performance and management initiatives were satisfactorily implemented, with involvement of agency senior leadership.

PROGRAM ASSESSMENT RATING TOOL

Perhaps the Bush administration's performance legacy will be best remembered for creation and implementation of the Program Assessment Rating Tool (PART). PART was the administration's response to a critique of GPRA implementation in preceding years, because the law only required collection of performance data without a mandate for using the information.[9] In response, the Bush administration designed PART as an accountability tool to publicly disseminate conclusions about program performance as a strategy for encouraging its use in decision-making. Thus, PART was intended to help fulfill the PMA's budget and performance-integration-reform goal by linking performance information with budget decisions.

The PART was developed by the Office of Management and Budget (OMB) in partnership with federal agencies—67 separate programs provided feedback about the instrument during a pilot phase.[10] The 25-question PART instrument emphasized program purpose and design, strategic planning, program management, and program results. Each section included questions that could be scored, then aggregated to generate a composite rating of program effectiveness. Not intended to itself serve as a formal program evaluation, half the composite index 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 PART rating process began with a negotiation between agencies and OMB on what constituted a program for purposes of the PART.[11] Agency coordinating staff, typically in budget offices, then worked with programs to respond to each PART question and develop supporting information to justify the response. Agencies submitted the self-assessment and supporting documentation to OMB career program examiners who then reviewed the documentation and finalized each PART rating. Once a program received a PART rating, scores were made publicly available on a website, ExpectMore.gov. In addition to the release of PART scores, brief public descriptions were provided, including a one-page cover note and a detailed description of how programs fared on each of the questions.

The Bush administration recognized that PART was an imperfect device for assessing program effectiveness. After a year of using the tool, OMB asked for feedback from agencies and made specific adjustments to address concerns raised.[12] The Government Accountability Office (GAO) identified more than 30 modifications in the instrument based on such feedback, including wording clarifications, additional or dropped questions, and merged questions.[13]

Within OMB, while individual program examiners were charged with conducting the reviews, the PART design was centrally managed by a dedicated staff, and later by the Division of Performance and Personnel Management. These offices of OMB career staff issued government-wide guidance, provided technical assistance to agency staff conducting assessments, developed the public website, ensured assessments were completed within given timeframes, and coordinated responses to agency appeals. The coordinating team's ability to articulate timelines and deadlines to agency and OMB staff, who had PART responsibilities layered on top of existing responsibilities, helped maintain the production timelines for PART reviews.

During the Bush Administration, over 1,000 programs received PART scores, including hundreds that were reviewed more than once.[14] By the end of the Administration, the PART process resulted in development of more than 6,000 performance metrics and more than 1,000 efficiency measures. But the proliferation of performance measures did not necessarily result in more evidence-based or evidence-informed policymaking.

Among the thousands of performance measures created, many were created solely for compliance purposes.[15] For instance, the Secret Service proposed outcome goals to achieve zero assassinations and injuries of protectees.[16] While these types of goals may be societally desirable, the existence of such goals in driving actual agent performance is questionable.

By executive order, President Bush created the Performance Improvement Council (PIC), a coordinating body to continue the PMA reform objectives by identifying senior agency leaders to coordinate performance management activities. PIC officials were encouraged to update agency strategic plans required under GPRA, ensure that aggressive program goals were set with reliable measurement procedures, routinely convene program staff to assess performance, and to incorporate performance measures in personnel appraisals.[17] Many agencies ultimately nominated their chief financial officers for the role, placing performance improvement activities across government largely within budgeting offices.

PROGRAM EVALUATION

Much of the Bush administration's emphasis on program evaluation was articulated through the PART. The tool included two specific questions on evaluation, to gauge whether programs had conducted summative or impact evaluations. The administration publicly called for "rigorous" evaluations to be requisites for federal funding.[18] In subsequent guidance issued by the OMB, experimental methods for evaluation were described as the premiere approach for providing such evidence for use within the executive branch.[19] The administration's guidance noted that experimental methods are not appropriate in all cases, while also discounting the value of other evaluation methods for assessing program activities. Inside OMB and some agencies, staff were trained by non-profit organizations to specifically promote experimental designs for evaluations.[20]

Some professional evaluators took issue with the pre-determined methodological approach in OMB's guidance and sought to engage and influence OMB to revise the guidance.[21] In addition, some professional evaluators further suggested the need for OMB to acknowledge the limitations of experimental methods and the relevance of mixed-methods approaches to address policymakers' questions.[22]

The Bush administration's PART process also launched a discussion within government about what constitutes an "independent" evaluation. Initial PART assessments only considered evaluations to be independent if they were conducted by inspectors general, GAO, or external evaluators with external funding. In subsequent years, the definition of independent evaluation was broadened to include agency evaluation offices and agency contract evaluations, and a range of methodologies.[23] Subsequent analyses of composite PART scores showed that programs that relied on internal evidence or qualitative evaluations tended to receive lower composite PART ratings, suggesting the presence of methodological preferences among OMB staff.[24]

Efforts to expand funding for evaluation activities met mixed results. While some programs with statutory evaluation directives and appropriations continued to produce evaluations, targeted new evaluations sought by the Bush administration faced some opposition in Congress. For example, Congress specifically prohibited a plan to conduct an evaluation of the Department of Education's Upward Bound program.[25] There were some notable changes, however, including the creation of the National Center for Education Evaluation and Regional Assistance in 2002 as part of the Education Sciences Reform Act.[26]

Several agencies also developed websites to disseminate agency-funded evaluations to encourage use. For example, the Department of Education's Institute for Education Sciences launched the What Works Clearinghouse (WWC) and the Substance Abuse and the Mental Health Services Administration launched the National Registry of Evidence-Based Programs and Practices (NREPP) to share results of completed studies publicly.[27] These types of actions to disseminate results from research and evaluation have been encouraged by professional associations and blue-ribbon panels to make evidence more accessible to potential users.[28]

OTHER INITIATIVES

Numerous other efforts beyond the evaluation and performance measurement activities pursued by the Bush administration occurred from 2001-2009, including major reforms for the statistical system. In 2002, the Confidential Information Protection and Efficiency Act (CIPSEA) was enacted, vastly strengthening the confidentiality protections available for government's statistical datasets and encouraging greater data sharing for certain economic data in the statistical system.[29] However, these efforts were not directly connected to other PMA or evaluation initiatives. Thus, the evidence initiatives largely did not consider how the 13 federal statistical agencies and numerous other statistical programs could help support performance measurement and evaluation activities.

Lessons from the Bush Administration's Initiatives

Reviews of the Bush administration's performance and evaluation strategies are now widely available, highlighting positive aspects, along with the more limiting features.[30] One general area of agreement arises: that the efforts led to increased compliance activities and drove programs toward achieved designated scores but not always meaningful improvement.[31]

The PART instrument sought to achieve multiple purposes: providing information for public and congressional accountability as well as informing congressional and executive funding decisions.[32] The diverse purposes limited the administration's ability to effectively target the tool to support both program improvement and accountability. Thus, presenting information to serve specific audiences or intended users could be a guiding principle for future reforms.[33] In addition, considering the balance between initiatives tailored to monitor performance and assess effectiveness simultaneously involves trade-offs that could be considered.[34]

PART was designed to provide public accountability by passing judgment on programs with the intent to invoke budgetary consequences. PART did enable public transparency of federal programs, but the provision of single composite PART scores and the categorical names for the ratings applied (e.g., "ineffective") contributed to apprehension among agency staff. The presumption that a single tool could dictate funding decisions must acknowledge the inherent subjectivity of performance information. For example, weaker than expected performance might produce different normative perspectives on funding needs. Thus, performance information is one of many factors decision-makers consider when determining program funding levels.

Federal agencies expended considerable resources implementing PMA reforms and conducting PART reviews during the Bush administration, yet congressional interest in performance information generated by the reviews was minimal.[35] Given the broader policy context, efforts to inform funding decisions with analysis derived from PART may not have gained traction in Congress due to lack of perceived need during strong economic conditions and relatively few budgetary pressures.

The Bush administration used the President's Management Council and PIC to engage senior agency officials, which impelled some agency participation and buy-in to the broader initiatives. By setting an expectation that performance was a priority and that agency leaders were accountable for demonstrating agency results, senior executives may have been more likely to engage in discussions on overall agency performance using provided information to guide decision-making. Though, combined with limited congressional interest, participation may have also been disincentivized for agency officials since it did not appear likely the PMA reports would affect agency budgets approved by Congress.[36]

PMA reforms required resource contributions from both agencies and OMB to implement. Completing a PART review with documentation and briefings for appropriate agency leaders and OMB typically required a dedicated coordinating staff. Given the increased workload associated with PMA activities, some agency staff tended to view the requirements as calls for compliance rather than performance enhancements.[37]

OMB served as the coordinating hub of the administration's PMA activities, including PART reviews, without which the PART would not likely have been implemented across government. At the same time, OMB's role and implementation strategy has been criticized for discrete issues. Some OMB staff inconsistently applied composite PART scores.[38] The PART required creation of new program efficiency measures, not all of which were necessarily useful for program management.[39] Some staff were critiqued due to a perceived to lack knowledge about certain program details.[40] The focus on uniformity with templates may have distracted from efforts to make sense of performance trends and the magnitude of problems faced to achieve meaningful improvements in program performance.[41] Taken together, these critiques suggest several challenges in building coalitions for successful implementation from a top-down model. However, the PMA reforms provided OMB staff leverage in engaging agency counterparts to discuss program performance, create new performance measures, and encourage evaluation production.

PART articulated the administration's prioritization of evaluation activities, and a view that measurement alone would not result in improved program performance. While stopping short of integrating evaluation and performance measurement activities in agencies, PART set the stage for OMB's framing of evaluation as a learning and improvement tool in subsequent years. In several identified instances, the mere presence of the questions about evaluation in the PART process encouraged programs to engage in new evaluation activities.[42]

However, narrowly defined requirements for evaluation "independence" and a methodological preference for experimental methods for evaluation activities created some tension for building greater evaluation capacity within the federal government and professional evaluation associations.

Obama Administration Performance and Evaluation Initiatives

The Obama administration described its interest in performance and evaluation as identifying a "common sense" approach for increasing government effectiveness.[43] President Obama 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."[44] The strategies employed by the Obama administration included developing an overarching management framework, replacing PART by instituting priority goals, and launching capacity strategies to facilitate greater use of impact evaluations.

MANAGEMENT AGENDA AND PRIORITY GOALS

The Obama administration's initial performance management agenda focused on addressing issues related to information technology, government real property, human capital, and customer service.[45] The initiative highlighted four areas: effectiveness, efficiency, economic growth, and people and culture.[46] The agenda involved monitoring through quarterly reviews, called "Stats," to discuss of performance data with senior officials to discuss progress toward achieving targets.

Much like the Bush administration's critiques of GPRA, the Obama administration viewed earlier performance reforms as unsuccessful in facilitating the intended use of performance information.[47] Within months of assuming office, the Obama administration announced the end of PART. The alternative approach to improving performance was to create high-level priority goals within and across agencies.[48] With development of the goals led by a new federal chief performance officer, agencies were directed to demonstrate progress toward the goals, explain progress, and hold managers accountable for goal attainment.[49]

"High-Priority" Performance Goals. In 2009, the Obama administration launched the development of what were initially called near-term, high priority performance goals for major agencies.[50] OMB directed agencies to: focus on priority goals with "unrelenting attention;" develop processes for sustained agency performance; work toward making GPRA documents more useful; and target challenges that would not likely be overcome in the absence of targeted resources and leadership attention.[51] Agencies were provided considerable flexibility and autonomy in specifying priorities, goals, and targets through an agency-driven process.[52] Part of the intentional approach from the Obama administration was to address the history of treating government's performance improvement initiatives as compliance activities.[53]

The Obama administration initially established more than 100 separate priority goals across federal agencies, and added hundreds more through its two terms. Some of the goals were modified over the years to reflect changing conditions and information. For example, initially the Department of Housing and Urban Development and the Department of Veterans Affairs jointly announced a goal of reducing the number of homeless veterans from about 200,000 to just 59,000 within two years.[54] Measured in 2014, the administration reported that less than 50,000 veterans were homeless, which resulted in a new goal being announced to *eliminate* veterans' homelessness altogether.[55] But, research on efforts to eliminate veterans' homelessness suggests that severe unintended consequences can result from what may appear as otherwise laudatory agency goals.[56] Like the transparency aims of PART, the Obama administration disseminated goals through a website on which performance could be monitored to facilitate accountability.[57]

GPRA Modernization. The enactment of the GPRA Modernization Act of 2010 codified much of the existing practice during the first two years of the Obama administration. The GPRA Modernization Act: (1) defined performance 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 priority 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 performance website.[58] GPRA Modernization also aligned the timeframes for strategic planning, performance reports, and goal setting requirements with the quadrennial election cycle and annual budget submissions, allowing plans and goals to better align with a president's priorities.[59]

In implementing GPRA Modernization, OMB outlined the objectives of the performance management framework—effectiveness, efficiency and productivity, transparency that engages the public, and fairness and equity.[60] Congress encouraged consultation on the new annual goals when developing the goals and making changes to priorities. OMB directed that annual goals be achievable within 24 months, solidifying their short-term design, and clarified that the priority goals previously developed would be "archived."[61] OMB staff's role in setting and reviewing targets for the annual goals was limited. Under GPRA Modernization, agencies are responsible for reviewing their progress on the various goals quarterly. However, for two other types of goals—cross agency and management goals—OMB was charged with setting the annual and quarterly targets, as well as convening relevant agencies quarterly to review progress. Over time the goal process became more institutionalized, though substantial challenges persisted in coordinating improvements across multiple agencies.[62]

EVIDENCE CAPACITY INITIATIVES

Early during Obama's first term, officials quickly announced an emphasis on improving program evaluation activities, although they avoided describing an integration with the priority goal efforts.[63] The administration's philosophy specifically noted the need for "rigorous, independent program evaluations," and asserted that evaluations were valuable for informing decisions to "find and fix or eliminate ineffective [programs]."[64] The administration emphasized using evaluation to assess causal impacts for determining whether to fund or not fund programs. By the second Obama term, senior policy officials from across the White House complex jointly suggested that evaluation is increasingly valuable during periods of constrained fiscal environments and to inform agencies' budget decisions.[65]

Much like the Bush approach, the Obama administration focused primarily on increasing the use of *impact* evaluation, asserting that existing evaluations had insufficiently influenced budget and management decisions, and that funds had been spent on evaluations of insufficient rigor or policy significance.[66]

The Obama administration's fiscal year 2010 budget proposed a modest initiative to improve evaluation transparency, promote evaluation with a staff workgroup, and provide additional funds for agencies that voluntarily committed to producing impact evaluations.[67] These early efforts segued into a wider focus on: (1) evaluation transparency; (2) training and collaboration; and (3) addressing evaluation barriers.

Evaluation Transparency and Use. In response to criticisms that government conducted evaluations that were never made publicly available, the administration committed to publicly listing evaluations funded by agencies to ensure results were publicly disseminated and to facilitate the eventual use of evaluation results. The administration announced an additional government-wide initiative to promote transparency, participation, and collaboration with the Open Government Directive.[68] The attempts to increase awareness for new or ongoing evaluations did lead to some successes. For example, the Department of Labor and the Department of Justice both launched new websites to disclose studies.[69] HHS's Administration for Children and Families institutionalized transparency by publishing an evaluation policy that called for transparency, independence, and policy relevance for all of the agency's evaluations.[70] The U.S. Agency for International Development produced similar guidance.[71] Other agencies like the State Department developed policies to promote internal dissemination of evaluations, and some limited public dissemination.[72]

Efforts to improve the public dissemination of evaluations resulted in development of a "common evidence framework."[73] The standards were agreed to by offices within HHS, Labor, and Education.[74] Some non-profit organizations also attempted to further the use of evaluation websites by creating techniques for encouraging development of clearinghouses and integrating websites and rating schemes into a single analytical tool.[75]

The administration promoted a tiered-evidence approach that incorporated evaluation requirements into competitive grant programs. The approach prioritized funding for grant proposals supported by existing causal studies, and created lower tiers for funding where evidence was in development so as to encourage continued programmatic innovation.[76] For example, the Department of Education's Invest in Innovation Fund (i3) scales funding based on the tier of evidence presented in a grant proposal.[77] One early assessment of the tiered-evidence strategy suggests that while the strategy is promising, it is too early to assess whether it was successful in deploying evaluation results to create a meaningful feedback loop for decision-making.[78]

Agency Collaboration and Leadership. The Obama administration created an interagency workgroup within the PIC to focus on strengthening agency capacity, sharing best practices, and providing technical assistance and research expertise to agencies.[79] Operationally, the efforts were limited in the scope and breadth of agency personnel involved. The group was eventually recast as a temporary series of forums and workshops for agency staff to share best practices and begin to identify techniques to integrate performance measurement and evaluation.[80] The Department of Education eventually took the lead in training efforts by establishing an Innovation Exchange, in coordinating with the PIC.[81]

The administration's efforts also emphasized the need to strengthen evaluation capacity, and encouraged policy officials to support the evaluation function. The administration encouraged agencies to consider options, such as the creation of chief evaluation officer positions, to empower agency officials to produce evaluations.[82] The Labor Department and the Centers for Disease Control created new chief evaluation officer positions.

OMB also launched the Interagency Council on Evaluation Policy, co-chaired by OMB staff and an agency evaluation leader. The organization brought together evaluation offices, including agencies without chief evaluation officers, to coordinate activities in a manner similar to the longstanding Interagency Council on Statistical Policy, which features the heads of the major federal government statistical agencies. One feature of the evaluation council's activities was to initiate a dialogue about principles and practices for evaluation in the federal government. In 2016, the National Academies of Sciences, Engineering and Medicine convened a panel to discuss principles and practices for evaluation units, with the goal of using the effort to bolster the integrity and objectivity of federal evaluations.[83]

Addressing Barriers to Evidence Building. During the Obama administration, OMB prioritized addressing three identified barriers to engaging in more evidence building about government programs, specifically the availability of resources, data collection limitations, and administrative data availability.

Enhancing Evaluation Funding. In 2009, the administration announced a $100 million evaluation initiative to provide new funding to 17 agencies to support specific evaluations and increased evaluation capacity.[84] While some of these proposals may have been funded by Congress, many evaluations in the proposal were ultimately not funded or undertaken by agencies. Some funding efforts were successful, such as the administration's request to reauthorize the demonstration authority for the Social Security Disability Insurance Program, which was reauthorized as part of the Bipartisan Budget Act of 2015.[85]

In later strategies the administration focused on increasing funding set-asides for existing programs, where a share of program resources could be devoted to evaluation. For example, multiple budgets requested authority for the Labor Department to allocate up to one percent of program funds to support research and evaluation, and 1.5 percent for HHS's Social Services Block Grant.[86]

The Obama administration also advocated the use of the Pay for Success contracting model as a way to encourage evaluation while limiting up-front government costs. Under the Pay for Success model, government contracts can be awarded based on local providers' ability to achieve specified performance targets, increasing payment when performance greatly exceeds targets and decreasing payment when performance goals are not achieved.[87] The Obama team advanced a strategy that supported the technique for use in government, non-profits, and the private sector.[88] However the approach largely avoided ongoing academic discourse about the model, its challenges, and risks for the mechanism in practice.[89]

The Obama administration pursued other efforts to improve funding flexibility for certain evidence-building activities. In 2015, the administration proposed to provide contracting flexibility to six departments to pilot streamlined procurement processes for evaluation that would have extended the availability of funding to accommodate evaluation-specific contracting processes and timelines, recognizing evaluations often occur over multiple years.[90] The authority was ultimately not supported by Congress.

Improving Data Collection Clearance Protocols. One potential impediment to federal evaluation long cited by evaluation practitioners is the formal clearance for any survey with ten or more participants required by the Paperwork Reduction Act.[91] The process is intended to limit the burden imposed on the public in responding to government data requests.[92] In 2009, OMB solicited public comments for improvements to how paperwork clearances were managed.[93] While the evaluation community viewed the request as an opportunity, no formal public action was taken by the administration to respond to received public comments.

Increasing Access to Government Data. A third barrier the Obama administration sought to address was limited access to existing administrative datasets.[94] Early efforts initiated access to non-confidential administrative data through websites like data.gov and for use in program evaluations and priority goal activities. Similarly, in 2014 the Digital Accountability and Transparency (DATA) Act was enacted with the goal of improving transparency for some government spending data.

Beyond the spending and data available through data.gov, the administration recognized that substantial barriers also persisted for confidential data, which include identifiable records. In 2014, OMB encouraged agencies to use their administrative data for statistical activities, including through coordination between statistical and program offices.[95] The administration also proposed targeted legal expansions of access to certain restricted datasets, such as the National Directory of New Hires among others, to be used for research and evaluation activities.[96] The Obama administration proposed and received additional funding with the Census Bureau in 2015 and 2016 to build technological infrastructure to improve access to state administrative datasets for federal programs and supported efforts to establish a commission that would recommend other ways to improve access to administrative data.[97]

Recognizing the need to develop a coordinated strategy to pursue improved access to administrative data, in 2016 Congress passed and President Obama signed a law creating the Commission on Evidence-Based Policymaking, an idea long-championed by Speaker Paul Ryan (R-WI) and Senator Patty Murray (D-WA).[98] While the commission did not offer its recommendations before the end of the Obama second term, the administration indicated optimism that the commission's approach would suggest strategies for making better use of government data and reducing barriers to their access.[99]

Rapid Cycle Evaluation. The Obama administration launched the Social and Behavioral Sciences Team in 2014, which would evolve into the Office of Evaluation Sciences at the General Services Administration, to encourage rapid cycle evaluation, though the projects implemented were limited in scope and scale. In 2015, the president signed an executive order directing federal agencies to apply behavioral insights in a manner to improve government services to the American public.[100] During the first year of the initiative, the office claimed to yield benefits in streamlining program access and improving program efficiency through small-scale projects at a variety of agencies focused on short-term outcomes.[101]

Lessons Learned from the Obama Administration Initiatives

In contrast to the Bush administration's initiatives, few assessments have been produced to date on the Obama initiatives. The following section reflects on what lessons were learned by executive branch staff relative to insights gained from the initiatives in the Obama administration.

The Obama priority goal effort was slow to launch, delayed by challenges in identifying and articulating the performance 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 evidence initiatives never materialized.

The administration promoted transparency in performance measurement and evaluation initiatives. However, the ability to translate the transparency efforts to public or congressional accountability was less clear. The Obama initiatives prioritized providing information about performance and results, but users needed the ability to meaningfully filter or translate that information.

With passage of GPRA Modernization and support for some early evaluation initiatives, there was a signal of interest from Congress in performance and evaluation not seen during the Bush Administration. The Obama team also continued to increase dissemination of performance and evaluation results on the internet. And priority goals limited the amount of performance information presented in open public forums, relegating other performance information to longer agency performance documents. But even then, the public mechanism for dissemination, Performance.gov, has been criticized for insufficiently addressing intended audiences, highlighting continued challenges in targeting information at the right level.[102]

Priority goals elevated the role of some senior leaders who were responsible for achieving targets and monitoring performance. In evaluation initiatives, new chief evaluation officer positions within a very few agencies established support for more and stronger evaluation. But senior officials were never required to utilize or pursue evaluation efforts, and in many policy areas never did. That said, senior White House officials were engaged early in the Obama administration in supporting the evidence agenda and helped push through some of the necessary statutory changes and funding with provisions favorable to the initiatives.[103] The high-level engagement by senior White House officials likely contributed to the initiative's endurance over the years.

The Obama initiatives returned the goal-setting focus to the agency level, and envisioned that individual political appointees would participate in the goal-setting to align interests with the president's priorities. While fewer key goals were developed and maintained in the Obama administration than under PART, those that were focused more on short-term achievements rather than long-term ones.

In addressing criticisms of the PART's uniformity, the Obama approach allowed agencies considerable flexibility in defining goals and targets. Within the evaluation initiatives, agencies were also given 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 more to social programs, despite potential gains from establishing evaluation cultures in other agencies and policy domains.

The structure of the new priority goals shifted workload away from OMB staff and onto agencies. Agencies led the effort to identify goals and to set targets for most of the priority 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 and negotiating final scores.

OMB's role under the Obama efforts was to design and globally coordinate the initiative, without interjecting to apply a heavy hand in setting goals. While OMB staff were more involved in coordinating and reviewing cross-agency goals, the emphasis on implementation and attainment continued to rest with agency staff. Similarly, for evaluation initiatives, OMB's role in the Obama administration was largely facilitative, encouraging agency staff to participate in various efforts.

While the Bush administration acknowledged a role for incorporating performance and evaluation within the PART instrument, the Obama administration did not seek integration of the disciplines. To some extent, the practices may be further apart today than at the beginning of the Obama term due, in part, to the creation of separate career staff offices for coordinating evaluation and performance measurement guidance within OMB. Even then, other efforts, like the Social and Behavioral Sciences Team, were established in yet other components of the Executive Office of the President. Statistical activities, which have potential to greatly benefit evaluation and performance offices are overseen by yet another office. The separation of activities mirrors the structure of many federal agencies which struggle to coordinate evidence-building activities.[104]

Lessons from Past Presidential Evidence Initiatives

The Bush and Obama administrations' efforts to improve performance measurement and evaluation systems were similar in approaches, yet both strategies yielded only modest results for encouraging the broader use of evidence in government. There are at least four overarching lessons that emerge from the strategies for future evidence-building activities: efforts need to be calibrated to support learning; efforts need coordination; audiences must be sustained; and effective implementation of the initiatives requires strategic alignment that engage the broader evidence-building community.

1: CALIBRATING EVIDENCE-BUILDING ACTIVITIES TO ACHIEVE LEARNING

Evaluation theory suggests there are two primary purposes for conducting evaluation: learning and accountability. Rarely does any initiative or data collection requirement satisfy both purposes. Yet, throughout the Bush and Obama administrations, while primarily motivated by the accountability purpose, the direction given to agencies espoused a desire for learning to improve performance.

Congress, OMB, and the American public are the primary audiences for developing information about government results, whereas users for performance and evaluation products tend to be agency leadership and program managers. But evaluation and performance measurement mandates are designed to benefit all audiences. Even the best-intentioned initiatives that are designed based on available theory inevitably motivate compliance with mandates, rather than providing real and lasting incentives to improve effectiveness. Initiatives can be better tailored to address the different needs of the intended audiences. In order to incentivize learning within agencies a more-resourced and targeted effort needs to be developed and implemented.

2: COORDINATING ACROSS THE EVIDENCE-BUILDING COMMUNITY

Current government practices largely do not integrate performance, evaluation, statistical, and research operations to support broader evidence-building goals and potential uses. Instead, evaluation and performance activities are often siloed with little coordination within agencies, and even within OMB. The various activities are viewed by agency program staff as separate enterprises, with little interaction, thus performance measurement and reporting do not benefit from statistical expertise or 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 oversight bodies like OMB and GAO. Addressing this challenge is vital, but quite arduous because bureaucratic inertia, cultural norms, and individual worldviews are difficult to modify. As the Commission on Evidence-Based Policymaking recommended in 2017, a convenient starting point for coordination would be for the OMB to better address breaking down operational silos to model for the rest of government how to approach coordinated evidence building.[105]

3: ACQUIRING AND SUSTAINING AUDIENCES OF POTENTIAL USERS

The establishment of demand for evaluation, met with sufficient supply, presents one strategy for developing and maintaining a use-case for performance information.[106] In contrast, a supply-driven approach would reflect instances where evaluations or performance information are produced without a clear end-user in mind (or the nebulous "everyone-is-a-user" argument), and are used when available. For either, there are real challenges considering consultation between the executive and legislative branches, and within the executive branch.

Even if consultation with congressional leaders had occurred in developing the Bush and Obama initiatives, little use of information provided to Congress appears to have occurred in practice. When Congress enacted the GPRA Modernization Act, congressional leaders could have created an optimal system for meeting congressional needs, though that did not occur. 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. Instead, presidential management and evidence agendas are just that—they are owned by the president and the executive branch, mostly the OMB and a few better-resourced agencies (e.g., Education, Health and Human Services, and Labor).

Neither the Bush nor Obama administration made notable strides in supporting political appointees to assume leadership and encourage visibility of the initiatives, though the assignment of priority goals to individuals within agencies was a productive start. With typically short tenures, political appointees have limited time to learn the ins-and-outs of programs' missions and then try to effectuate performance improvements. Performance or evaluation initiatives are typically not among new political appointees' highest priorities. And there has not been any visible nor sustained training for political appointees on evidence-building capacity.

Of note, neither administration developed and disseminated sufficient success stories to relay to executive leaders, Congress, or the public that the performance or evaluation initiatives "worked." When stories were shared, they were frequently the same few stories, perceived to be generous representations of reality, leading agency staff to become skeptical of the initiatives' value. Because leadership for performance or evaluation initiatives relies heavily upon persuasion, sharing of success stories can be beneficial.

4: IMPLEMENTING INITIATIVES EFFECTIVELY

Even the best designed, theory-based initiative must face the realities of implementation in complex, federal institutions. Theory in the private sector has suggested that the act of setting goals can improve government performance in terms of both effectiveness and efficiency. But in practice, the perverse incentives introduced with mandating performance and evaluation activities may lead implementers to focus on the measurable rather than the desirable. For example, in the veterans' homelessness goal cited above, subgroup analysis was limited due to measurement constraints; therefore, the incentives established for measuring homelessness resulted in a subpopulation realizing devastating consequences.[107]

Similarly, all of the initiatives previously described struggled to appropriately calibrate 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 ability and influence to ensure agency staff are attentive to prescribed requirements. But when agency staff perceive that 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 tend to 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 modeling and process evaluations. In such cases, agencies will almost always have difficulty "owning" new initiatives.

Implementing government-wide, or cross-agency initiatives effectively to improve government performance has been a repeated challenge, even with OMB leadership in coordination of the initiatives. 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 fiscal year 2016 budget, requesting and eventually receiving a small amount of funding to help coordinate cross-agency priority goals, but agencies need more incentives and resources to fully collaborate on complex issues that span intergovernmental jurisdictions to be effective.[108] Similarly, building evaluation capacity in agencies beyond Education, Labor, and HHS is necessary to support evidence-building initiatives over the long-term. For example, the social and behavioral sciences initiative was undertaken as a consulting effort and was not perceived as being owned by key agency managers. Basic capacity can help agencies identify which data to collect, and to appropriately analyze the data to learn important lessons about programs. Numerous initiatives in both administrations alienated agencies by calling for the use of randomized controlled trials in impact evaluation. Some agencies' programs and expertise are simply not well-suited to pursue these approaches.

Conclusions

For the next era of evidence-*informed* policymaking, efforts to incorporate information into decision-making must recognize that performance metrics and evaluation are not the only inputs into the complex decision-making apparatus. Since the 1990s a prevalent idea has been that policymakers are capable of identifying needed information and gauging the appropriateness of applying such information to inform policy decisions. But determining what constitutes valid, relevant, and reliable evidence for many agencies is complicated for program staff—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 performance measurement and evaluation activities remains limited, if not inconsistent, across federal agencies, and perhaps more so across states. The federal government could continue to explore strategies to better develop capacity to align with the dual purposes of accountability and learning. Efforts to develop baseline assessments of agency capacity may offer a starting point for central coordinating and oversight offices to prioritize resources for program improvement. This was part of the motivation for the Commission on Evidence-Based Policymaking's 2017 recommendation to inventory offices that engage in evidence-building activities to better understand what resources can be brought to bear.[109]

Future initiatives may also benefit by focusing on the theoretical interrelationship between performance and evaluation activities that has proven evasive in practice, as well as improving relationships between all evidence-building functions. In monitoring and evaluation systems, indicators of program performance can highlight projects meeting defined targets, provide indicators for impact assessments, and fulfill overarching accountability requirements.[110] Rarely are performance monitoring and measurement envisioned as activities that occur independently from program management. Routine measurement and evaluation can positively reinforce each other, and evaluation can become more useful and less costly when programs have repeatedly collected data and refined metrics to improve performance. Efforts to realize this interconnectedness may, in turn, help improve the quality of evidence, while reducing the costs of developing evidence.

In practice, disconnects between the evaluation and performance measurement functions have persisted across administrations and through countless initiatives, though so too has the disjointed nature of the broader evidence-building community. Current practice within the federal government tends to emphasize performance measurement systems, where practitioners operate in silos, are too often assigned to different organizational units, and are not viewed as owned by program managers. The Bush administration designed PART to provide an overarching performance and analytical framework, although the tool emphasized performance measures more than evaluation. Similarly, the Obama administration espoused the benefits of integrating the efforts, even while adopting a management framework that largely continued the bifurcation. While the theoretical importance for pursuing integration is evident, the practical challenges of integrating monitoring and evaluation activities are numerous. One path forward might be to encourage agencies to combine the functional positions and offices within agencies that oversee evaluation and performance efforts. 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.[111]

Finally, administrations appear to quickly recognize it is one thing to announce new initiatives and another to implement them well. While some lessons are important irrespective of strategy, the value of others rests on the vision for use. In addition, the vision must strike an appropriate balance between uniformity and flexibility, and centralization and decentralization. But above all, the value of engagement for senior agency leaders—both political and career—and external stakeholders is paramount to fulfilling any initiative that seeks to improve decision-making. Without buy-in from senior leaders who perceive information as unbiased, valid, and useful, and who are willing to utilize the information developed through performance and evaluation efforts, achieving the initiatives' success will almost always be futile.

Both the Bush and Obama performance measurement and evaluation strategies provided lessons for future initiatives. This technical paper highlighted many similarities across the administrations, some differences, and lessons to be extracted from both. As the Commission on Evidence-Based Policymaking described: "To achieve the greatest gains, evidence-building must be well-coordinated both within and across departments."[112] The evidence initiatives over the past 20 years suggest that future evidence-building initiatives aimed at improving government's effectiveness can certainly be better designed to integrate performance measurement and evaluation activities with the numerous other efforts within the evidence-building community. Doing so will not only improve how efforts to build evidence occur, but also better inform the variety of potential uses among potential users.

References

1   Osborne, David and Ted Gaebler. (1992). *Reinventing Government: How the Entreprenuerla Spirit is Transforming the Public Sector*. Addison-Wesley.

2   Gore, Al. (1998). *Common Sense Government: Works Better and Costs Less: National Performance Review (3rd Report)*. Washington, D.C.: National Performance Review.

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

4   Joyce, Philip. (2011). The Obama Administration and PBB: Building on the Legacy of Federal Performance-Informed Budgeting? *Public Administration Review*, 71(3), 356-367.

5   U.S. House of Representatives. (1995) Report 104-99. Available at: https://www.congress.gov/104/crpt/hrpt99/CRPT-104hrpt99.pdf

6   Gore, Al. (1997). *Businesslike government: lessons learned from America's best companies*. Washington, D.C.: National Performance Review.

7   Kettl, Donald. (1998). *Reinventing Government: A Fifth-year Report Card*. Washington, D.C.: The Brookings Institution.

8   Office of Management and Budget (OMB). (2001). *The President's Management Agenda*. Washington, D.C.: OMB.

9   OMB. (2003). *Budget of the United States Government, Fiscal Year 2004: Performance and Management Assessments*. Washington, D.C.: GPO.

10   *Ibid*.

11   Joyce, 2011.

12   OMB, 2003.

13   Government Accountability Office (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.

14   U.S. (2008). *ExpectMore.gov*. Washington, D.C. http://georgewbush-whitehouse.archives.gov/omb/expectmore

15   Moynihan, Donald. (2013). Advancing the Empirical Study of Performance Management: What We Learned From the Program Assessment Rating Tool. *American Review of Public Administration, 43*(5), 499-517.

16   U.S., 2008.

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

18   OMB, 2001.

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

20   Haskins, Ron and Greg Margolis. (2014). *Show Me the Evidence: Obama's Fight for Rigor and Results in Social Policy*. Washington, D.C.: Brookings Institution Press.

21   Bernholz, Eric, et al. (2006). *Evaluation Dialogue Between OMB Staff and Federal Evaluators*. Washington, D.C.

22   Trochim, William. (2008). *Response to OMB re: What Constitutes Strong Evidence of a Program's Effectiveness*. Available at: http://www.eval.org/d/do/98

23   GAO, 2004.

24   Heinrich, Carolyn. (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.

25   United States (U.S.). (2007). Consolidated Appropriations Act, 2008, Public Law 110–161, Section 519 (December 26). Available at: https://www.gpo.gov/fdsys/pkg/PLAW-110publ161/pdf/PLAW-110publ161.pdf

26   U.S. (2002a). Education Sciences Reform, Public Law 107-279 (November 5). Available at: https://www.gpo.gov/fdsys/pkg/PLAW-107publ279/pdf/PLAW-107publ279.pdf

27   Institute for Education Sciences (IES). (2014). What Works Clearinghouse. Available at: http://ies.ed.gov/ncee/wwc/ and Substance Abuse and Mental Health Services Admnistration (SAMHSA). (2014). National Registry of Evidence-Based Programs and Practices. Available at: http://www.nrepp.samhsa.gov.

28   American Evaluation Association (AEA). (2013). Evaluation Roadmap for a More Effective Government. Washington, D.C.: AEA. Available at http://www.eval.org/d/do/472

29   U.S. (2002b). E-Government Act of 2002, Public Law 107-347. Available at: https://www.gpo.gov/fdsys/pkg/PLAW-107publ347/html/PLAW-107publ347.htm

30   See Heinrich, 2012; Joyce, 2011; Moynihan, 2013; Moynihan, Donald and Stephane Lavertu. (2012). Does involvement in performance management routines encourage performance information use? Evaluating GPRA and PART. *Public Administration Review, 72*(4), 592-602; and White, Joseph. (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.

31   Metzenbaum, Shelly and Robert J. Shea. (2017). Performance Accountability, Evidence and Improvement: Bipartisan Reflections and Recommendations to the Next Administration. *PA Times*, pp. 18-25.

32   Newcomer, Kathryn and F. Steven Redburn. (2008). Achieving Real Improvement in Federal Policy and Program Outcomes. *Obama-Biden Transition Project*. Washington, D.C.: National Academy of Public Administration and George Washington University.

33   Metzenbaum, Shelley. (2009). *Performance Management Recommendations for the New Administration*. Boston: Center for Public Management.

34   Newcomer and Redburn, 2008.

35   Breul, Jonathan, and John Kamensky. (2008). Federal Government Reform: Lessons from Clinton's "Reinventing Government" and Bush's "Management Agenda" Initiatives. *Public Administration Review, 68*(6): 1009-1026.

36   Newcomer & Redburn, 2008.

37   *Ibid. – Newcomer and Redburn*

38   Moynihan, 2013.

39   Metzenbaum, 2009.

40   White, 2012.

41   Metzenbaum, 2009.

42   Hart, N. (2016). Evaluation at EPA: Determinants of the U.S. Environmental Protection Agency's Capacity to Supply Program Evaluation. Diss. Washington, D.C.: George Washington University.

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

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

45   OMB. (2010a). *The Accountable Government Initiative—An Update on Our Performance Management Agenda.* Washington, D.C.: OMB.

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

47   OMB. (2009a). Building a High-Performing Government *Analytical Perspectives: Budget of the United States Government for Fiscal Year 2010.* Washington, D.C.: GPO.; OMB. (2010b). Delivering High-Performance Government Performance and Management *Analytical Perspectives: Budget of the United States Government for Fiscal Year 2011*. Washington, D.C.: GPO.

48   *Ibid*.

49   Brass, Clint. (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.

50   OMB. (2009e). *Planning for the President's Fiscal Year 2011 Budget and Performance Plans (M-09-20)*. Washington, D.C.: OMB.

51   OMB. (2010c). *Performance Improvement Guidance: Management Responsibilities and Goverment Performance Results Act Documents (M-10-24)*. Washington, D.C.: OMB.

52   Joyce, 2011.

53   Metzenbaum and Shea, 2017.

54   OMB. (2011c). *High Priority Performance Goals*. Washington, D.C.: OMB. Available at: http://www.whitehouse.gov/sites/default/files/omb/performance/high-priority-performance-goals.pdf

55 U.S., 2015.

56 Augeri, Justine. (2015). *Supportive Services for Homeless Veteran Women*. Dissertation. Washington, D.C.: George Washington University.

57 OMB, 2010c.

58 OMB. (2011b). *Delivering on the Accountable Government Initiative and Implementing the GPRA Modernization Act of 2010 (M-11-17)*. Washington, D.C.: OMB.

59 Brass, 2012.

60 OMB. (2011a). *Delivering an Efficient, Effective, and Accountable Government (M-11-31)*. Washington, D.C.: OMB.

61 OMB, 2011b.

62 OMB, 2015.

63 OMB, 2009a; OMB. (2009b). *Increased Emphasis on Program Evaluations (M-10-01)*. Washington, D.C.: OMB; Orszag, Peter. (2009). Building Rigorous Evidence to Drive Policy. Available at: https://obamawhitehouse.archives.gov/omb/blog/09/06/08/BuildingRigorousEvidencetoDrivePolicy/.

64 OMB, 2009b; OMB. (2011e). Program Evaluation *Analytical Perspectives: Budget of the United States Government for Fiscal Year 2012*. Washington, D.C.: GPO.

65 OMB, OSTP, DPC, & CEA. (2013). *Next Steps in the Evidence and Innovation Agenda (M-13-17)*. Washington, D.C.: White House.

66 OMB, 2009b

67 *Ibid*.

68 OMB. (2009d). *Open Government Directive (M-10-06)*. Washington, D.C.: OMB.

69 Department of Labor (DOL). (2014a). *Clearinghouse for Labor Evaluation and Research*. Washington, D.C. Available at: http://clear.dol.gov/; National Institute for Justice. (2014). CrimeSolutions.gov. Available at: https://www.crimesolutions.gov

70 HHS. (2014). *Evaluation Policy*. Federal Register, Washington D.C.: GPO.
    Available at: https://www.federalregister.gov/articles/2014/08/29/2014-20616/evaluation-policy-cooperative-research-or-demonstration-projects

71 USAID. (2016). USAID Evaluation Policy. Available at:https://www.usaid.gov/sites/default/files/documents/1870/USAIDEvaluationPolicy.pdf

72 Department of State. (2015). Department of State Evaluation Policy. Available at:https://www.state.gov/s/d/rm/rls/evaluation/2015/236970.htm

73 OMB. (2013). *Budget of the United States Government, Fiscal Year 2014*. Washinton, D.C.: GPO.

74 HHS. (2013). Common Evidence Framework (Draft). Washington, D.C.
    http://findyouthinfo.gov/docs/Common_Evidence_Framework-Draft_508_5-08-13.pdf ; DOL. (2014b). *CLEAR Causal Evidence Guidelines*. Washington, D.C. Available at: http://clear.dol.gov/sites/default/files/CLEAR_EvidenceGuidelines_1.1_revised.pdf ; and IES and the National Science Foundation. (2013). Common Guidelines for Education Research and Development. Available at: https://ies.ed.gov/pdf/CommonGuidelines.pdf

75 America Achieves. (2013). Investing in What Works Index. Available at: http://www.americaachieves.org/docs/InvestInWhatWorksIndex.September2013.pdf; Pew. (2014). Results First Clearinghouse Database.
    Available at: http://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2014/09/results-first-clearinghouse-database

76 OMB. (2010d). Program Evaluation *Analytical Perspectives: Budget of the United States Government for Fiscal Year 2011.* Washington, D.C.: GPO.

77 *Ibid*; and Donovan, Shaun. (2017). OMB Cabinet Exit Memorandum.
    Available at: https://obamawhitehouse.archives.gov/sites/default/files/omb/reports/cabinet_exit_memorandum.pdf

78 Haskins and Margolis, 2014.

79 OMB, 2009b.

80 OMB et al., 2013.

81 Department of Education. (2016). *Innovation Exchange*. Washington, D.C. Available at http://www2.ed.gov/about/offices/list/ods/evidence/

82 OMB et al., 2013.

83 National Academy of Sciences, Engineering, and Medicine (NASEM). 2017. Principles and Practices for Federal Program Evaluation Proceedings of a Workshop—in Brief. Washington, D.C.
Available at: https://www.nap.edu/catalog/24716/principles-and-practices-for-federal-program-evaluation-proceedings-of-a

84 OMB, 2009b; 2010d

85 U.S. (2015). Bipartisan Budget Act of 2015, Public Law 114-74. Available at: https://www.gpo.gov/fdsys/pkg/PLAW-114publ74/pdf/PLAW-114publ74.pdf

86 OMB, 2015; OMB. (2016a). *Budget of the United States Government, Fiscal Year 2017*. Washington, D.C.: GPO.

87 Liebman, Jeffrey. (2011). Testing Pay-for-Success Bonds. *Public Manager, 40*(3): 66-68.

88 OMB. (2011d). *Paying for Success*. Washington, D.C.: OMB.
Available at: https://obamawhitehouse.archives.gov/administration/eop/sicp/initiatives/pay-for-success

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

90 HHS. (2015). 2016 Budget Request: Administration for Children and Families. https://www.acf.hhs.gov/sites/default/files/olab/2016_acf_cj.pdf; OMB, 2015.

91 Rog, Debra, William Trochim and Leslise Cooksy. (2009b). *Improving Implementation of the Paperwork Reduction Act.*
Available at: http://www.eval.org/d/do/104; U.S. Commission on Evidence-Based Policymaking (CEP). (2017). The Promise of Evidence-Based Policymaking: Final Report of the Commission on Evidence-Based Policymaking. Washington, D.C.

92 OMB. (2009c). *Information Collection under the Paperwork Reduction Act* Washington, D.C.: OMB.

93 74 Federal Register 206: 55269.

94 OMB. (2016b). Barriers to Using Administrative Data for Evidence-Building. Washington, D.C.: OMB.
Available at: https://obamawhitehouse.archives.gov/sites/default/files/omb/mgmt-gpra/barriers_to_using_administrative_data_for_evidence_building.pdf

95 OMB. 2014. M-14-06: https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf

96 NASEM. 2017. Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy. Washington, D.C. https://www.nap.edu/catalog/24652/innovations-in-federal-statistics-combining-data-sources-while-protecting-privacy

97 OMB, 2015, 2016a.

98 U.S. (2016). Evidence-Based Policymaking Commission Act, Public Law 114-140.
Available at: https://www.gpo.gov/fdsys/pkg/PLAW-114publ140/pdf/PLAW-114publ140.pdf

99 Donovan, 2017.

100 Obama, Barack. (2015). *Executive Order -- Using Behavioral Science Insights to Better Serve the American People*. Washington, D.C.: White House.

101 Office of Science Technology & Policy (OSTP). (2016). *Social and Behavioral Sciences Team Annual Report*. Washington, D.C: White House; and Shankar, Maya. (2014). *How low-cost randomized controlled trials can drive effective social spending*. Washington, DC: White House Office of Science and Technology Policy.
Available at: https://obamawhitehouse.archives.gov/blog/2014/07/30/how-low-cost-randomized-controlled-trials-can-drive-effective-social-spending

102 GAO. (2013). *Managing For Results: Agencies Should More Fully Develop Priority Goals under the GPRA Modernization Act.* Washington, D.C.: GAO.

103 Haskins and Margolis, 2014.

104 CEP, 2017.

105 *Ibid.*

106 Hart, 2016.

107 Augeri, 2015.

108 OMB, 2015

109 CEP, 2017.

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

[111] Newcomer, Kathryn and Clint Brass. (2016). Forging a Strategic and Comprehensive Approach to Evaluation within Public and Nonprofit Organizations: Integrating Measurement and Analytics within Evaluation. *American Journal of Evaluation 37*(1): 80-99.

[112] CEP, 2017, p. 101.

BIPARTISAN POLICY CENTER

The Bipartisan Policy Center is a non-profit organization that combines the best ideas from both parties to promote health, security, and opportunity for all Americans. BPC drives principled and politically viable policy solutions through the power of rigorous analysis, painstaking negotiation, and aggressive advocacy.

bipartisanpolicy.org | 202-204-2400
1225 Eye Street NW, Suite 1000
Washington, D.C. 20005

@BPC_Bipartisan
facebook.com/BipartisanPolicyCenter
instagram.com/BPC_Bipartisan

BPC POLICY AREAS

Economy

Education

Energy

Evidence

Finance

Governance

Health

Housing

Immigration

Infrastructure

National Security

BIPARTISAN POLICY CENTER