

Putting it all together: The case of the U.S. Department of Labor's evidence-building strategy

Molly Irwin¹ | Demetra Smith Nightingale² 

¹PEW Charitable Trusts, Philadelphia, Pennsylvania, USA

²Urban Institute, North Andover, Massachusetts, USA

Correspondence

Molly Irwin, PEW Charitable Trusts, Philadelphia, PA, USA.

Email: molly.e.irwin@gmail.com

Abstract

This chapter describes how and why the U.S. Department of Labor (DOL) structured and implemented a comprehensive evidence-building strategy in the years ahead of the federal legislation that now requires many of the same key components. In 2010, the Chief Evaluation Office was established in DOL at the departmental level to coordinate evaluation strategy and evidence building and to promote an organization-wide culture of learning. This represented a new approach intended to elevate the priority on evidence, improve the scope and quality of evaluations and research, and expand the use of evidence. The DOL strategy included formalizing a departmental evaluation policy statement around key principles that govern high-quality evaluations, developing a learning agenda process to strategically plan for evaluations and evidence-building activities, and creating an evidence-based clearinghouse to synthesize and share the results of rigorous evaluations. While each department is unique, DOL's experience highlights functions that were prioritized as well as challenges and limitations that had to be addressed in one department.

The Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act) directs federal agencies to strengthen and institutionalize evaluation practice in various ways. The Act includes a number of recommendations presented in 2017 by the U.S. Commission on Evidence-Based Policymaking, and more generally builds on growing bipartisan attention to and interest in the importance of creating and supporting an evidence infrastructure in the federal government (CEP, 2017). Several departments, such as the U.S. Department of Labor (DOL), the U.S. Department of Education, and the Administration for Children and

The authors thank the Data Foundation for funding for open access.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *New Directions for Evaluation* published by American Evaluation Association and Wiley Periodicals LLC.

Families at the U.S. Department of Health and Human Services (HHS), already had evaluation, research, and evidence-based structures. The procedures and operational details varied across agencies, each with approaches that had developed and evolved over several decades (Haskins & Margolis, 2014; Maynard et al., 2016).

While attention to evidence-based policy started before 2018, the Evidence Act added momentum, responding to growing interest for a more systematic approach to data, research, and evaluation, and created a mandate to build evidence capacity throughout the federal government. Federal evaluation capacity and activity was, and to some extent still is, uneven across agencies. A U.S. Government Accountability Office (GAO) report in 2014, for instance, found that fewer than half of agencies reported committing resources to evaluation or having a central office responsible for evaluation, and only half of these offices were reported to have a stable source of funding. Furthermore, the GAO report found that just seven agencies reported having a high-level official responsible for evaluation. In addition, only a quarter of agencies reported having agency-wide policies or guidance concerning key issues such as study design, evaluator independence and objectivity, report transparency, or implementing findings. The agencies and subagencies with evaluation and research offices, though, reported greater evaluation coverage and more use of the results in decision-making (U.S. Government Accountability Office, 2014).

The Evidence Act includes several provisions that require many federal agencies to establish or strengthen evidence practices and activities. All departments subject to the Chief Financial Officer provisions are required to put into place a number of practices, such as:

- Designating an evaluation officer and office
- Coordinating evaluation, data, and statistical functions
- Having an evaluation policy statement
- Producing and publishing learning agendas and evaluation plans

Guidance from the Office of Management and Budget strongly encourages departments beyond those subject to the CFO provisions to also employ these practices. Guidance and implementation timelines for these activities are outlined in OMB (2019, 2020, 2021).

DOL embarked on a strategic approach to building evidence in 2011 and put in place many of these practices prior to the passage of the Evidence Act. This experience serves as a useful case study about both opportunities and challenges in implementing practices now required across the federal government. This chapter describes evidence structures and activities in DOL and contextual factors that helped or hindered progress under two presidential administrations, four labor secretaries, and numerous changes to laws, agency leadership, and staff.

ESTABLISHING THE EVALUATION OFFICE AND ITS ROLE

In 2010, DOL established the Chief Evaluation Office (CEO) to encourage and build the capacity for, and the understanding and use of, evaluation and other types of evidence throughout the department. The first CEO was appointed and a small staff with research and evaluation expertise and experience was assigned to the office to coordinate, manage, and implement the department's evaluation approach.

The office at DOL was established to serve a number of key evidence building and evidence use functions. The primary motivation for having the CEO—a departmental level evaluation office—was to design, fund, and oversee the implementation of rigorous, high-quality evaluations aimed at building evidence about programs and policies throughout

the department. Additionally, establishing an executive position to oversee this function was a way to promote transparency and accountability. Projects were aimed at answering a wide range of questions and included various methodological approaches, for example: outcome and impact evaluations (both experimental and nonexperimental, and including behavioral economics studies), implementation analysis, feasibility studies, exploratory studies, systems change studies, studies aimed at understanding and synthesizing the state of the evidence in a particular area (e.g., literature reviews, evidence reviews, meta-analysis), and studies aimed at filling data and information needs, such as statistical analysis and surveys.

The number of projects the CEO funds and oversees has grown over time, ebbing and flowing with fluctuations in the budget. Through most of its tenure (and currently), the CEO has 50 or more active projects at any given time. These studies are primarily carried out by independent external evaluators through contracts and, in some cases, grants. The CEO also works with subagencies within the department to undertake evaluations and research that are not funded or directly overseen by the CEO. The CEO had a small staff, fewer than twenty evaluation specialists in 2020, augmented by a few contractors and a couple of staff on temporary assignment from other offices. The CEO staff also work closely with staff from research and evaluation units that exist in several subagencies, such as the Employment and Training Administration, the Employee Benefits Security Administration, the Occupational Safety and Health Administration and others, which have staff dedicated to research, analysis, and evaluation.

In addition to funding, overseeing, and collaborating on evaluations, the CEO also plays a major role in improving access to administrative and program data for analysis and providing support and technical assistance on analytic methods. To this end, in 2013 the CEO was charged by DOL's acting secretary with creating a Data Analytics Unit to support and complement efforts to improve the use of administrative data. This unit was established to provide quantitative data analysis and analytic consultation to DOL subagencies and programs to facilitate the use of existing data for a variety of analyses, research, and evaluations. Some analysis may be for internal program purposes, such as linking business data across labor standards enforcement programs. Data analysts also conduct special studies, such as analyzing Bureau of Labor Statistics data to prepare profiles of workers in specific sectors (Schafer et al., 2015). They also collaborate with other staff in the CEO or in program offices to facilitate the use of administrative or other existing data for evaluations, such as improving the interfaces between workforce development and veterans employment program reporting systems. In 2019, the head of the Data Analytics Unit became the department's Chief Data Officer required by the Evidence Act. He remains co-located with the CEO and continues to work alongside the data analytics team and the evaluation specialists in the CEO, continuing the close alignment of evaluation and data.

COLLABORATING ACROSS TYPES OF EVIDENCE

There are many types of evidence in addition to program evaluation and data analysis. Each department is unique in how involved an evaluation office might be with other evidence activities. At DOL, the CEO is primarily responsible for program evaluations and related research including data analytics support, as just discussed, but the CEO coordinates closely with other evidence-building activities, especially performance measurement and statistical activities.

Program evaluation and performance measurement are closely related but have unique purposes and analytic methods (Hatry, 2013). There are different definitions for

evaluation, but at the most general level, it involves assessing situations, conditions, or issues. In OMB guidance, evaluation is defined as “an assessment using systematic data collection and analysis of one or more programs, policies, and organizations intended to assess their effectiveness and efficiency” (2020, p.1). At DOL, as in many federal departments, evaluation draws mainly from scientific and social scientific theory and methodology to systematically assess the design, implementation and outcomes of a policy or program and identify ways to improve results (Rossi et al., 1999; Shadish et al., 1990). The studies are typically conducted periodically, usually by outside independent researchers, and apply a range of quantitative and qualitative analytic methods.

Performance measurement, in contrast, is an internal administrative function intended to allow managers and staff to continuously record and measure activities and progress over time. Ideally, a performance measurement system is designed to set priorities and goals and support tracking progress toward specifically defined metrics and targets (Harris, 2015). Performance measurement draws mainly from management science and public administration, and usually uses internally produced data. It is a continuous activity ideally providing a feedback loop for accountability and program improvement. Performance management in the federal government includes complying with requirements set forth in the Government Performance and Results Act (GPRA). GPRA was amended in the GPRA Modernization Act (GPRMA) of 2010. The GPRMA requires, for example, systematic quarterly reviews of all measures by each department’s Chief Operating Officer (Newcomer & Brass, 2021).

Beginning in 2009, DOL began reforming the departmental performance management process to better align with the purpose of GPRA and GPRMA. This involved each operating subagency and program in DOL reviewing and prioritizing measures in a quarterly meeting with the Deputy Secretary, and sometimes the Secretary, to review progress, challenges, administrative data quality issues, and operational or budgetary implications (Harris, 2015). In 2014, each subagency had an average of between 10 and 20 priority measures, which were aligned with the department’s 4-year strategic plan.

To reinforce the importance of both evaluation and performance, the Chief Evaluation Officer was included in all quarterly performance meetings. In addition to reviewing the quarterly and annual performance metrics, recently completed, current, and potential new evaluations were often discussed. The goal was to consider findings from evaluations along with the performance measures to help administrators understand the body of data and evidence around a particular program or issue. For example, evaluations of the Job Corps residential training and services program for disadvantaged youth suggested that the outcomes for older youth were greater than outcomes for youth under 18 years of age. The evaluation findings along with performance and management evidence were all used to develop new targeted strategies for the younger participants.

Including the Chief Evaluation Officer in the quarterly performance review meetings helped improve administrators’ understanding of what evaluation is and how it differs from, but is complementary with, performance measurement. In time, some Assistant Secretaries routinely asked to meet with the CEO ahead of the quarterly meetings to receive status reports on ongoing projects or discuss research ideas. Including evaluation in the meetings also served to solidify the partnerships needed to address critical issues and challenges around the quality of administrative data, accessing program and statistical data for evaluation and research, and including evaluations and resulting findings in operating plans and in the department’s strategic plan.

At DOL, another important type of evidence comes from the analysis of statistical information and series administered by the Bureau of Labor Statistics (BLS). While coordination between evaluation and statistics was not initially a priority when the CEO was

established, over time a very close interaction developed, again reinforced by the quarterly performance meeting process. In the Bureau of Labor Statistics quarterly performance meetings, discussions ranged from issues and challenges arising with specific surveys and data, resource constraints around data systems, and sharing and comparing results from statistical analysis and evaluations. Additionally, the learning agenda process (discussed below) often highlights evidence gaps identified by the subagencies that could be filled by joint work. Thus, in subsequent years the BLS and CEO partnered to initiate, test, analyze, or supplement some critical surveys and data series, such as a Survey of Employer Provided Training and analysis of the Time Use Survey data and special analysis of the first responder workforce.

In setting up the CEO in 2010, coordination across types of evidence was a priority, and aligning program evaluation and performance measurement was an important and conscious activity. The cross-evidence coordination also figured importantly in implementing the Evidence Act. This can be seen in DOL's data governance framework (DOL, 2019) to coordinate the key evidence-building roles outlined in the Act. The Secretary's Order mandates collaboration among the Chief Data Officer, the Chief Performance Officer, Chief Evaluation Officer, Chief Information Officer, and Chief Statistical Officer. It stipulates a collaborative approach to reviewing IT infrastructure and data asset accessibility, developing solutions for managing, disseminating and generating data, coordinating statistical functions, supporting evaluation, research and evidence generation, and supporting all aspects of performance management including assurances that data are fit for purpose.

STRUCTURING THE EVALUATION OFFICE

Establishing the CEO and its role was not without challenges. Structural change is not easy. At first, there were no dedicated resources for the CEO and negligible funding for evaluation projects. In the initial two years of the CEO, support and funding were dependent on funds from the Offices of the Secretary and Deputy Secretary, meaning the existence of the office was somewhat fragile, as were decisions about the best organizational location for the office. While there was some consideration of locating the evaluation office with either the performance management center or the Bureau of Labor Statistics, neither option was chosen, in part because the activities are distinct and in part because there was no dedicated funding for the office. For administrative purposes, the office was placed in the Office of the Assistant Secretary for Policy, but as an independent unit with the Chief Evaluation Officer position lateral to a deputy assistant secretary.

The evaluation office was gradually formalized through the federal budget process. In 2012, the department proposed and obtained Congressionally appropriated dedicated annual funding for evaluations and the office. Since that time, Congress has specifically appropriated funds to be used for program evaluation under the direction of the Chief Evaluation Officer. Starting in 2013, the department also received budget set-aside authority for evaluations, allowing the Secretary to set aside up to 0.5% (later increased to 0.75%) of operations funds in the department for evaluations.¹ Other departments, such as the U.S. Department of Agriculture's Food and Nutrition Service and the Administration for Children and Families in HHS have also set-aside funding for evaluation and research, with their own specific features. Set-aside funds at DOL are resources that are appropriated for specific operational or programmatic activities that can be transferred to the evaluation office and used for evaluation activities; that is, other funding can be set aside.

The DOL evaluation budget set-aside is one of the most important assets for the CEO, but also presents challenges because of its uncertainty and budgetary tradeoff

considerations. The DOL set-aside authority provides the Secretary with the discretion to set the percentage up to the maximum allowed each year. During the administration of President Barack Obama, the Secretary approved the CEO plan to set aside approximately \$30 million a year for evaluation activities developed in consultation with operating agencies (still far under the maximum amount allowed by law. The Labor Department's annual discretionary budget is approximately \$12 billion). However, during the administration of President Donald Trump, the set-aside budget decreased by about 90% (Data Coalition, 2019). A firm minimum to ensure some resources are set aside, as well as maximum to retain secretarial discretion, would provide certainty and allow for more strategic evaluation planning. Nonetheless, the CEO appropriations and set-aside funding allowed the office to sponsor evaluations, maintain a small core staff of evaluation specialists and invest in key capacity-building activities such as establishing a formal evaluation policy statement, creating a learning agenda and evaluation planning process, and developing an evidence-based clearinghouse.

As the evaluation office at DOL grew, the leadership arrangement was modified, to ultimately support permanency. Initially, to set up the office quickly and develop and refine activities and processes, the first chief evaluation officers were appointed under special personnel statuses. Once the office was established, the position became a career position under the senior executive service (SES). When the CEO was being set up, the first Chief Evaluation Officer, Jean Grossman, served in non-career status while on leave from her academic position at Princeton University. The second officer, Demetra Nightingale, was appointed for two years under the Intergovernmental Personnel Act authority and later served as a senior non-career appointee. After 5 years, the Chief Evaluation Officer position was established as a senior career position and in 2016 Molly Irwin became the first career Chief Evaluation Officer at DOL to be competitively selected. The flexible, but short-term, appointments in the early phases helped speed the implementation of the office, and the eventual career status helped institutionalize and provide continuity for the position and the office.

DEVELOPING AN EVALUATION POLICY STATEMENT

To reinforce the emphasis on evidence-building and the role of the evaluation office in DOL, a Departmental Evaluation Policy statement was developed in 2011 (DOL, 2011). The policy presents the principles that guide DOL's planning, conduct, and use of program evaluations. It emphasizes a commitment to conducting rigorous, relevant evaluations and to using evidence from evaluations and research to inform policy and practice.

The DOL evaluation policy statement drew heavily from one already in place in the Administration for Children and Families at the Department of Health and Human Services, one of the long-standing evaluation offices in the federal government. By 2017, several agencies had evaluation policies, each with similar principles, but defined in ways that aligned with their unique missions and programs (Nightingale & Scott, 2018). DOL's evaluation policy includes the following core principles:

- **Rigor:** commitment to using the most rigorous methods appropriate and feasible within statutory, budget, and other constraints including using the highest quality study designs, methodologies, analyses, and reporting methods.
- **Relevance:** addressing issues, questions, and topics of high interest to the agency, program, stakeholders, Congress, and the current administration; and producing results that can be used to help inform improvement in programs, services, outcomes, and impacts.

- **Transparency:** commitment to make information about evaluations and findings from evaluations broadly available and accessible by making studies and reports available within agencies and to the public, identifying the evaluator, releasing study plans, and describing the evaluation methods.
- **Independence:** ensuring objective and unbiased studies and researchers, and after technical peer review, the evaluation officer has authority to approve, release, and disseminate evaluation reports.
- **Ethics:** following professional research and evaluation standards; protecting privacy and confidentiality of study participants and human subjects.

The principles included in DOL's and other agencies' evaluation policy statements appear in Exhibit 1 (Nightingale & Scott, 2018).

The evaluation policy statement at DOL, like the one at the Administration for Children and Families, was intended to operationalize and institutionalize an evidence culture throughout the agency. The policy is built around the five principles noted above, to infuse evaluation into agency operations and coordinate research and evaluation with performance management, strategic planning systems, and statistical data activities.

The evaluation policy helped the CEO create a language and framework around evidence-building in the department. The policy and the principles articulate a vision, a common goal, and some key processes. The framework and concepts were useful in creating productive relationships with program and policy partners in the offices and operating subagencies in DOL.

However, building trusting relationships and managing the inherent tension that can exist between program and evaluation activities required consistent attention. First, developing a formal policy or operational statement in a federal department requires considerable time and involves building support and consensus about the principles and the wording. The statement went through full departmental clearance, which included legal review as well as review, comment, and agreement from the heads of all subagencies. At DOL, the formal clearance and revision process took about four months. The effort was worthwhile, though, because once finalized and agreed upon by DOL's senior leaders, it served as a commitment for all research and evaluation in the department, not just studies initiated by the CEO.

The part of the statement that involved the most discussion and drew some concern was independence, which is fundamental to the basic principles of professional evaluation standards (National Academy of Sciences, 2017; Yarbrough, Shula, Hopson, & Caruthers, 2010). During the full departmental review, some subagency leaders raised issues about when and who could release final evaluation reports. It is understandable that both political and career agency leaders are often apprehensive about the possible negative results of evaluations and research. For example, one hears anecdotally of instances throughout the government where administrators or political appointees may delay or restrict distribution of final evaluation reports or other research with scientifically obtained findings. There are some legitimate reasons to restrict reports. At DOL for instance, the labor standards enforcement programs involve legal and regulatory investigation, and judicial action. After consulting with other departments with enforcement responsibilities (e.g., Internal Revenue Service, Department of Justice), the DOL evaluation policy statement language was revised. Studies that are primarily for internal management, procedural or judicial purposes, such as randomized tests of alternative case review procedures, can be designated "for internal use only," but that determination must be made before a study begins, not once the findings are known. That is, all reports and findings are publicly released, regardless of whether the findings are positive or not, unless legal or judicial restrictions are

EXHIBIT 1 Examples of federal agency evaluation policy principles (in 2018)**U.S. Department of Health and Human Services,
Administration for Children and Families
(HHS-ACF)**

- **Rigor:** Using the most rigorous methods appropriate to the evaluation questions and feasible within budget and other constraints.
- **Relevance:** Evaluation priorities should take into account legislative requirements and Congressional interests and should reflect the interests and needs of ACF, HHS, and Administration leadership; program office staff and leadership; ACF partners such as states, territories, tribes, and local grantees; the populations served; researchers; and other stakeholders. Evaluations should be designed to represent the diverse populations that ACF programs serve, and encourage diversity among those carrying out the work.
- **Transparency:** Make information about planned and ongoing evaluations easily accessible. Release results in a timely manner regardless of the findings. Evaluation reports will describe the methods used, strengths and weaknesses, and the generalizability of findings. Evaluation reports will present comprehensive results, including favorable, unfavorable, and null findings.
- **Independence**—Insulate evaluation functions from undue influence and from both the appearance and the reality of bias. To promote objectivity, ACF protects independence in the design, conduct and analysis of evaluations. Evaluations are conducted through the competitive award of grants and contracts to external experts who are free from conflicts of interest. The director of the Office of Planning, Research and Evaluation reports directly to the Assistant Secretary for Children and Families; has authority to approve the design of evaluation projects and analysis plans; and has authority to approve, release and disseminate evaluation reports.
- **Ethics:** Evaluations will be conducted in an ethical manner and safeguard the dignity, rights, safety, and privacy of participants. ACF-sponsored evaluations will comply. ACF-sponsored evaluations will be conducted in an ethical manner and safeguard the dignity, rights, safety, and privacy of participants. ACF-sponsored evaluations will comply with both the spirit and the letter of relevant requirements such as regulations governing research involving human subjects.

**U.S. Agency for International Development
(USAID)**Evaluations should be

- **Integrated into design** of projects, strategies, and activities.
- **Unbiased** in measurement and reporting.
- **Relevant** to future decisions.
- Oriented toward reinforcing local **ownership**, consistent with institutional aims of local ownership through respectful engagement with all partners, including local beneficiaries and stakeholders, while leveraging and building local evaluation capacity.
- **Transparent**—shared widely with commitment to full and active disclosure.

(Continues)

EXHIBIT 1 (Continued)**Small Business Administration (SBA)**

- **Ethics:** Conduct the evaluation by adhering to the rules governing human rights, confidentiality, and privacy. Minimize the burden to research participants and cost to taxpayers.
- **Independence:** Conduct the evaluation through an outside party that does not have vested interest in the outcome or will not interpret the results in ways that are self-serving or misleading. Eliminate the appearance of bias to ensure results are properly used.
- **Rigor:** Employ the methodological approaches that best support the definitive answers to the evaluation questions under investigation
- **Relevance:** Scope and select evaluation questions most closely tied to the goals of the program, the priorities of the Agency, and the intended use by senior leaders.
- **Transparency:** Ensure that the evaluation, scope, design, implementation, and results are available for internal and public review, assessment, and critique.

U.S. Department of Labor (DOL)

- **Rigor:** Use the most rigorous methods appropriate and feasible within statutory, budget, and other constraints. Evaluation staff are skilled in the methods.
- **Relevance:** Evaluation priorities take into account legislative requirements and the interests and needs of leadership, specific agencies, and programs; program office staff and leadership; states, territories, tribes, grantees, populations served, researchers, and other stakeholders.
- **Transparency:** Information about evaluations and findings from evaluations are available and accessible. Release in a timely manner results of all evaluations that are not specifically focused on internal management, legal, or enforcement procedures or that are not otherwise prohibited from disclosure. Make evaluation data available as public use for secondary analysis, with appropriate privacy and data security.
- **Independence:** Insulate evaluation functions from undue influence and from both the appearance and the reality of bias. After technical peer review, the Chief Evaluation Officer has authority to approve, release, and disseminate evaluation reports.
- **Ethics:** Evaluations will be conducted in an ethical manner and safeguard the dignity, rights, safety, and privacy of participants and human subjects, complying with both the spirit and the letter of relevant laws and regulations.

Source: Nightingale & Scott, 2018

identified upfront (between 2011 and 2017, only two such studies were carried out by the DOL CEO, both involving legal processes).

USING A LEARNING AGENDA AND EVALUATION PLANNING PROCESS

Soon after establishing the CEO, evaluation specialists began working with each operating subagency to create five-year learning agendas. The learning agenda process is implemented in collaboration with program administrators. Each learning agenda highlights priority questions of interest to the program, subagency or department and identifies data and research needed to address those questions. The subagency agendas are then used by the CEO to develop the annual evaluation plan that identifies particular projects. The main intent is to initiate evaluations to help improve the efficiency and effectiveness of programs by building the evidence base around topics of interest to programs.

The learning agenda process is an important planning tool at DOL. The learning agenda is not viewed just as a document to be produced nor as the final evaluation plan, but as a *process*. The concept of the learning agenda as a process is also evident in OMB guidance (OMB, 2021; Newcomer et al., Chapter 5). Each year (and often multiple times a year), the CEO staff meet with administrators and staff in the operating agencies, and often with program offices within agencies. They discuss the status of evidence in a specific

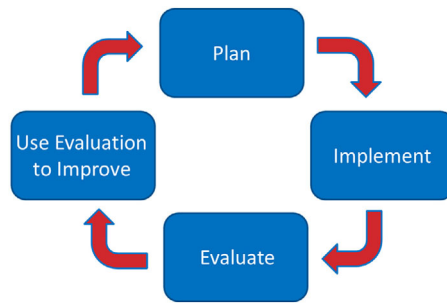


FIGURE 1 Evidence-building cycle

programmatic area, review the interim or final results of evaluations, and engage in conversations about the operational and policy implications of findings. Often staff in operating programs may identify additional information or knowledge they would like to improve their decision-making or improve programs.

These types of meetings and conversations often surface key questions and priorities for future studies that would be relevant and useful to the agency and other stakeholders. They also serve as an opportunity to share information about how to incorporate evaluation, research findings, and evidence into programmatic decision-making. Thus, the process can and should help develop an evidence-building cycle, as shown in Figure 1, as well as promote and support evidence use and ensure that the studies executed are important and relevant to the agencies and programs.

Finally, and importantly, the conversations that take place during this process should help to build trust and collaborative relationships between those operating programs (those being evaluated) and the CEO staff responsible for planning and overseeing evaluations. When that happens successfully, the learning agenda process can help to create demand and support from operating agencies for the services of the evaluation office. This is particularly important when the agencies, through the set-aside, are helping to fund the studies.

In addition to engaging internal stakeholders, the learning agenda process also provides an opportunity to engage external stakeholders, including field operations staff, service delivery operators, grantees, researchers, evaluators, advocates, and other officials at all levels of government. The Evidence Act guidance from OMB suggests the importance of stakeholder engagement in evidence-building activities (Epstein et al., Chapter 6). When the CEO was first established, the primary stakeholders and “clients” were internal—the operating agencies, departmental and agency leadership, and the staff and programs in states and localities. Over time, the CEO has broadened stakeholder input. The evaluation plan, for example, is published in the *Federal Register* and specifically requests review and comment from interested parties (Fed Reg, 2016). Going beyond the *Federal Register* notice, the CEO also seeks external input on special evaluation issues and topics. For example, several of the large research and policy topics on which portfolios of studies are designed (e.g., policies and programs for Native Americans and for youth disconnected from school and work) have convened workgroups that include representatives of affected groups or areas and issued formal Requests for Information.

As the CEO expands its external stakeholder input to learning agendas, they are drawing upon several decades of experience in the Employment and Training Administration (ETA), the largest subagency in DOL. In developing the five-year evaluation plan required by the Workforce Innovation and Opportunity Act, ETA research staff hold conference calls and request written input from several dozen researchers, program administrators, and

association representatives to identify priority issues on which future research or evaluations should focus. This input, as well as input from national and regional office staff, is incorporated into the ETA learning agenda process with the CEO. It seems safe to assume that including external input will likely continue to expand in DOL.

The CEO staff draft a department-wide evaluation plan based on the subagency learning agendas, the department's strategic plan priorities, statutory requirements for evaluations, and secretarial and administration policy priorities. The learning agenda is a 5-year roadmap and the annual evaluation plan describes the studies and activities that will be undertaken in the next year. The CEO finalizes the department's annual evaluation plan in close coordination with the operating agencies, and incorporates comments received on the draft from outside the department, for example from OMB or from public comments. The plan identifies priority evaluation and research studies and evaluation-related activities (e.g., literature reviews, data analysis, syntheses) building upon learning agendas. The plan indicates specific evaluations and related activities that can be funded with direct appropriations to the CEO and set-aside funding transferred to the CEO from other offices in DOL.

The evaluation set-aside provision in the annual appropriations bills requires the Chief Evaluation Officer to notify the U.S. Senate and House Committees on Appropriations about planned evaluations that will be carried out using the set-aside authority at least 15 days before the funds are transferred to the CEO. The annual DOL evaluation plan serves as the Congressional notification.

There are always more research questions or possible studies than there are resources (capacity and/or budget). Thus, the evaluation plan prioritizes studies that agencies have indicated are of high importance or that focus on measuring the efficiency and effectiveness of program outputs and outcomes consistent with the Secretary's priorities, the departmental strategic plan, and the agency operating plans. Finally, consistent with the DOL evaluation policy, the CEO encourages and prioritizes the most rigorous evaluation designs of programs and strategies possible, using the highest social scientific and evaluation methodological standards.

DISSEMINATING RESULTS AND USING EVIDENCE

As indicated in the evidence-building cycle depicted in Figure 1, the main purpose of evaluations and research is to inform decision-making and improve results. Each year, literally thousands of research projects are initiated or completed by, or on behalf of, one or more federal departments. If the results of studies do not reach users, this important purpose cannot be met (GAO, 2013).

The role of social science, including evaluation in public policy, has been a topic of research by political scientists and sociologists for decades (Nathan, 1988; O'Connor, 2001; Tseng, 2009; Weiss, 1979). Since the late 1990s and at the turn of the twenty-first century, there has been increasing attention to elevating the role of evidence, including promoting rigorous methodological analysis, and understanding the use of evidence in public policy. The federal government and philanthropic foundations are among those promoting the importance of the use of evidence and research. The William T. Grant Foundation, for example, has funded numerous grants to build and test theories of evidence use, and support research-practice partnerships to include practitioner interests in evaluations (Tseng, 2017).

The use of evidence has also been emphasized in OMB guidelines (2019, 2020, 2021) as well as in recent years as part of the President's Budget request to Congress, usually with

a separate section in the Analytic Perspectives volume. In a section of the Analytical Perspectives chapter of the Fiscal Year 2018 Budget (entitled, Building and Using Evidence to Improve Government Effectiveness), OMB stated that, "Agencies should integrate quality evidence and rigorous evaluation into budget, management, and policy decisions through a broad set of activities" (OMB, 2017, p. 55).

Using evaluation findings and research evidence, of course, requires that they are made available to stakeholders to improve programs, as well as to peer researchers who may review or replicate design, methodologies, and findings. Therefore, the CEO places high priority on making research reports and results readily available to the public, policymakers, researchers, and program operators, and embeds conversations about evidence use into its practice. At DOL, as noted earlier, implications of research and evaluations are discussed with stakeholders in quarterly performance meetings, during the learning agenda process, and when designing new studies. Reports are publicly released, and broadly disseminated, to encourage use.

In the past decade, calls for using evidence in decision-making have expanded. Congress often includes evaluation and evidence requirements in laws affecting federal agencies, including DOL. For example, the legislation authorizing the Reemployment Services and Employability Assessment Program for recipients of unemployment insurance requires every state that receives funding to implement strategies on which there is evidence of a positive impact.² If there is no evidence, the program must be evaluated.

To strengthen dissemination and evidence use, in 2011, DOL embarked on developing an evidence-based clearinghouse—the Clearinghouse for Labor Evaluation and Research (CLEAR). CLEAR drew on lessons from the Department of Education's What Works Clearinghouse (WWC), but was designed to meet the particular missions of DOL. It is intended to cover all subagencies and programs in DOL—those with large bodies of evaluations and research as well as those that were just beginning to conduct formal evaluations. All relevant studies are included, not just those funded by DOL. Similar to WWC and other clearinghouses, CLEAR presents ratings of the quality of rigorous evaluations. Eventually, CLEAR also published guidelines for conducting high-quality evaluations using various methodologies, began to include ratings of effectiveness based on the findings, and produced syntheses and short summaries of evidence for program operators.

Clearinghouses can play an important role in conveying findings about the body of evidence that may be useful for policy or programmatic decision making, and to determine where and how to strengthen the evidence base and build a research agenda. A main challenge in developing CLEAR was how to design a clearinghouse that was useful, accessible, and relevant for a broad group of stakeholders and end users, including those in DOL and across its subagencies. The potential users vary in terms of policy interest, and include stakeholders inside DOL and those outside, such as researchers and evaluators.

Importantly in the evidence-based context, the subagencies have different types of experience sponsoring evaluations and research and in using research evidence, whether sponsored by the department or not. Some offices have evaluation and research units and fund projects with grants or contracts; others have very little experience with research aside, perhaps, from management or monitoring studies. The Employment and Training Administration has perhaps the most evaluation experience and also maintains two research repositories separate from CLEAR. While they do not rate methodological rigor, they both have considerable detail and information useful for program operators: a research archive of all DOL-funded studies on ETA programs and a separate website with summaries of study findings, regardless of funding source. The ETA repositories have been designed specifically for program operators in the field. CLEAR includes many studies that are in the ETA systems but also rates the quality of methodologies to give users an

indication of how much confidence they should have in a particular study and how strong the evidence base is across a topic area.

In addition to making CLEAR useful for all of the DOL, another equally critical challenge regards resources. Developing an evidence-based clearinghouse, with methodological guidelines, rating standards, platform designs and interfaces, and the hundreds of systematic reviews that must be done is costly. There was preliminary agreement with the Department of Education that DOL could adapt the What Works Clearinghouse platform, but once design work began it became obvious that this would not work, in large part because of the wide range of policy and program areas of interest to DOL and because DOL's system was to include various types of evaluations, not just impact studies using experimental and nonexperimental designs.

At DOL, having an evidence-based clearinghouse was considered a worthwhile investment because of the importance of making findings readily available to the entire labor system. It was also a high priority for the CEO, with important collaboration from the Chief Economist and the Deputy Secretary. However, there was no separately appropriated funding for CLEAR, so resources from the CEO's core budget were dedicated on an annual basis to continue to develop and refine the system. After a 2-year development period, the system became operational, and it continues to be refined and improved each year. CLEAR is used in various ways, such as synthesizing results of multiple evaluations in nontechnical forms. It is also used to identify evidence-based strategies in particular areas such as reemployment services for unemployed workers or employment services for veterans, and ways to promote voluntary business compliance with labor standards requirements. CLEAR is also used in grant-making. The CLEAR standards for rigorous methodologies are referenced in announcements of program grant funding available when a condition of the award is that the program be rigorously evaluated. Over time, CLEAR's standards, ratings, and syntheses should improve the quality of evaluations being conducted as well as the use of the findings to improve programs and services.

CONCLUSION

The evaluation office at DOL provides one example of how a federal department expanded its focus on producing and using high-quality evidence and coordinating different types of evidence. Establishing the office benefited from support from multiple DOL Secretaries and Deputy Secretaries, OMB evidence and evaluation guidelines, and the collaboration between CEO and program office managers and staff. A number of challenges had to be addressed along the way, but the CEO has evolved over more than a decade and continues to refine the ways it carries out evidence-related activities: operationalizing an evaluation policy statement, using a learning agenda and evaluation planning process to set study priorities, collaborating with other offices in the department on evidence functions, and disseminating evaluation results in a way that can they can be useful to a broad range of internal and external stakeholders.

A few summary points are offered here:

First, establishing and building a successful federal evaluation office may not be easy, cheap, or fast. But not everything has to be done at once. At DOL, a few priority tasks were undertaken simultaneously to set a firm foundation for (slowly) building a culture of evidence department-wide and increasing the proportion of studies that met rigorous methodological standards. The three early priorities were: (1) drafting and obtaining full departmental approval of an evaluation policy statement, (2) putting

a learning agenda process in place, and (3) developing an evidence-based clearinghouse. Once these were operationalized, other activities became more feasible.

Second, there is no one way for every federal agency to proceed. Each department is unique, and the processes and structures needed to build a culture of evidence or learning should be tailored to the organization and its mission. DOL is a mid-sized federal department, with under 15,000 employees and nearly two dozen subagencies and offices. The CEO is a departmental level chief evaluation office with planning and data responsibilities as well as a comprehensive agenda of evaluations funded by contracts and grants. In a much larger department, a different type of evaluation office may be more appropriate—such as one that serves a more coordinating or assistance role.

Third, even though every department is different, DOL benefited greatly from the experiences and lessons from other departments. For example, the general framework for CLEAR benefited from the Department of Education's What Works Clearinghouse, and DOL's evaluation policy statement drew heavily on the one from the Administration for Children and Families at HHS. This suggests knowledge sharing and a cohesive community of practice for new evaluation officers in government, which has been developing since the passage of the Evidence Act, may offer substantial support to the capabilities and capacity in many departments.

Finally, it is critical for an evaluation office to remain relevant to the core mission(s) of the department. This includes maintaining close, productive, and collaborative relationships with programs on which evaluations may focus and with political and career leaders and administrators whose support can help maintain the office. Evaluators and researchers must understand that different users and stakeholders will value evaluation through a variety of perspectives, meaning technical reports may benefit from more concise briefs or summaries focused on program and policy implications, in addition to the technical reports of more interest to peer evaluators. Practices at DOL, such as developing an evaluation policy and learning agenda process, were designed to build a culture of evidence and to be positive and collaborative to allay program concerns or fears about evaluations. The CEO strives to improve familiarity with evaluation practices and build a culture that values and uses evidence and views study findings—positive or negative—as an opportunity for learning and improving program design and outcomes. Including the internal stakeholders in evidence-based activities is essential. Engaging them in developing a shared set of principles, centering learning agendas and evaluation plans on their decision-making needs, and providing support and tools such as an evidence clearinghouse to answer the question relevant to them, can help gain their support and trust in building evidence capacity in the department.

ORCID

Demetra Smith Nightingale  <https://orcid.org/0000-0002-4896-8024>

ENDNOTE

¹The annual appropriations bills specify the set-aside to be applied to specific budget accounts. For example, the Consolidated Appropriations Act, 2021 (P.L. 116-260, pp. 382-383) states, "The Secretary may reserve not more than 0.75 percent from each appropriation made available in this Act identified in subsection (b) in order to carry out evaluations of any of the programs or activities that are funded under such accounts. Any funds reserved under this section shall be transferred to "Departmental Management" for use by the Office of the Chief Evaluation Officer within the Department of Labor, and shall be available for obligation through September 30, 2022: *Provided*, That such funds shall only be available if the Chief Evaluation Officer of the Department of Labor submits a plan to the Committees on Appropriations of the House of Representatives and the Senate describing the evaluations to be carried out 15 days in advance of any transfer."

²The Bipartisan Budget Act of 2018, Public Law 115–123 (BBA) created a permanent authorization for the RESEA program and included evidence-based programming requirements. The RESEA provisions are contained in Section 30206 of the BBA, enacting new Section 306 of the Social Security Act.

REFERENCES

- Commission on Evidence-Based Policymaking (CEP). (2017). *The promise of evidence-based policymaking: Report of the commission on evidence-based policymaking*. GPO.
- Data Coalition. (2019). *Data coalition obtains documents indicating labor department is constraining resources for evaluating program effectiveness*. Retrieved from <https://www.datacoalition.org/press-releases/data-coalition-obtains-documents-indicating-labor-department-is-constraining-resources-for-evaluating-program-effectiveness/>
- Federal Register (2016) *Request for information; Comment request*, Department of Labor Research and Evaluation Plan for 2016, Notice by the Department of Labor, Jan 8 2016.
- Harris, S. D. (2015). Managing for social change: Improving labor department performance in a Partisan Era. *West Virginia Law Review*, 117, 987–1046.
- 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 [Special Issue: Performance Management and Evaluation]. *New Directions for Evaluation*, 2013(137), 19–32.
- Maynard, R., Goldstein, N., & Nightingale, D. (2016) Program and policy evaluations in practice: Highlights from the federal perspective [Special Issue in Social Experiments in Practice]. *New Directions for Evaluation*, 2016(152), 109–135.
- Nathan, R. P. (1988). *Social science in government: Uses and misuses*. Basic Books.
- National Academies of Sciences, Engineering, and Medicine. (2017). *Principles and practices for federal program evaluation: Proceedings of a workshop*. The National Academies Press. <https://doi.org/10.17226/24831>
- Newcomer, K., & Brass, C. (2021). Making sense of where we've been and what's next, the public productivity and performance. *The public productivity and performance handbook* (3rd ed., pp. 43–68, Ch. 2). In Marc Holzer & Andrew Ballard (Eds.), Routledge.
- Nightingale, D. S. & Scott, M. (2018). *Building evidence culture and capacity in federal agencies*. Urban Institute.
- O'Connor (2001). *Poverty knowledge: Social science, social policy, and the poor in twentieth-century U.S. history*. Princeton University Press.
- Office of Management & Budget, Exec. Office of the President (2017). Analytical perspectives budget of the U.S. government fiscal year 2018. Chapter 6: Building and using evidence to improve government effectiveness (pp. 55–58). Retrieved from <https://www.govinfo.gov/content/pkg/BUDGET-2018-PER/pdf/BUDGET-2018-PER.pdf>
- Office of Management & Budget, Exec. Office of the President (2019). *OMB M-19-23, Phase 1 implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning agendas, personnel, and planning guidance*. Retrieved from <https://www.whitehouse.gov/wp-content/uploads/2019/07/M-19-23.pdf>;
- Office of Management & Budget, Exec. Office of the President (2020). *OMB M-20-12, Phase 4 implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program evaluation standards and practices*. Retrieved from <https://www.whitehouse.gov/wp-content/uploads/2020/03/M-20-12.pdf>;
- Office of Management & Budget, Exec. Office of the President (2021). *OMB M-21-17, Evidence-based policymaking: Learning agendas and annual evaluation plans*.
- Rossi, P., Freeman, H., & Lipsey, M. (1999). *Evaluation: A systematic approach* (6th ed.). Sage Publications.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1990). *Foundations of program evaluation: Theories of practice*. SAGE Publications.
- Schafer, K., Sutter, R., & Gibbons, S. (2015). *Characteristics of individuals and employment among first responders*. U.S. DOL, Chief Evaluation Office.
- Tseng, V. (2009). *Studying the use of research evidence in policy and practice*. William T. Grant Foundation.
- Tseng, V. (2017) The next big leap for research-practice partnerships: Building and testing theories to improve research use. *The Digest*, Issue 3, Winter.
- U.S. Department of Labor. (2011). *U.S. Department of Labor Evaluation Policy*. Retrieved from <https://www.dol.gov/agencies/oasp/evaluation/EvaluationPolicy>
- U.S. Department of Labor. (2019). *Secretary's Order 02-2019-Chief Data Officer and DOL Data Board*. 84 FR 11327 (03/26/2019). Retrieved from <https://www.federalregister.gov/documents/2019/03/26/2019-05720/secretarys-order-02-2019-chief-data-officer-and-dol-data-board>
- U.S. Government Accountability Office (2013) *Program evaluation: Strategies to facilitate agencies' use of evaluation in program management and policy making*. GAO-13-570.

U.S. Government Accountability Office (2014). *Some agencies reported that networking, hiring, and involving program staff help build capacity*. GAO-15-25: Nov 2014. Retrieved from <http://www.gao.gov/products/GAO-15-25>

Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review*, 39(5), 426–431.

Yarbrough, D. B., Shula, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Corwin Press.

AUTHOR BIOGRAPHIES

Molly Irwin is the Vice President for Research and Science at the Pew Charitable Trusts. She was the Chief Evaluation Officer at the U.S. Department of Labor from 2016 to 2019.

Demetra Smith Nightingale is an Institute Fellow at the Urban Institute. She was the Chief Evaluation Officer at the U.S. Department of Labor from 2011 to 2016.

How to cite this article: Irwin, M., & Nightingale, D. S. (2022). Putting it all together: The case of the U.S. Department of Labor's evidence-building strategy. *New Directions for Evaluation*, 2022, 101–116. <https://doi.org/10.1002/ev.20489>