URBAN INSTITUTE · ELEVATE · THE · DEBATE

Making Evidence Relevant to Government: The Role of Evaluators and Researchers

Demetra Smith Nightingale

November 2017

Evidence-based decisionmaking has gained considerable traction over the past decade in the federal government. The term "evidence" is used in many different ways, but fundamentally it means using information that is produced from a systematic analysis of results. In public policy, evidence means analysis of the results of public programs, policies, or strategies. Public policy researchers and program evaluators play a critical role because they conduct the empirical analysis that produces the evidence. Using research tools and methods standard in various academic disciplines and fields, evaluators and researchers measure program outcomes, analyze factors that underlie outcomes, estimate impacts and costs of programs and strategies, and disaggregate factors that contribute to trends and results. Most researchers hope that policymakers and decisionmakers use the findings that emerge from their studies, but there is no guarantee.

In fact, it is not unusual for research reports to "sit on shelves" untouched and unread. Articles in scholarly journals may receive attention from other researchers and academics, but few reach a policy practitioner. Even short policy or practice papers (often referred to as "briefs"), written specifically to draw out the program and policy implications of a research study or evaluation, may not adequately reach or be relevant to the policy or program audience.

There are several ways researchers and evaluators can ensure that their studies are as relevant and useful as possible to agencies and administrators who are responsible for the programs being evaluated. The following thoughts are based on my experience as a researcher and, most recently, as chief evaluation officer at the US Department of Labor.

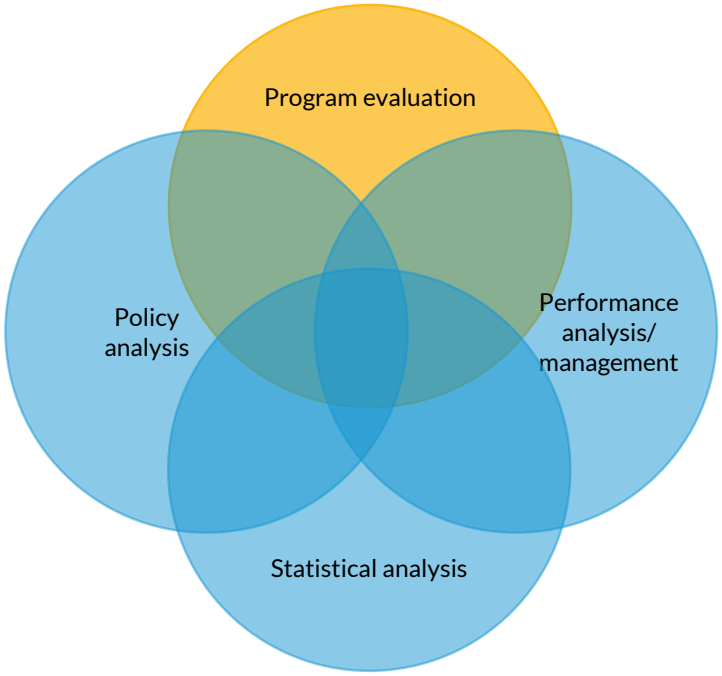
Recognize That There Are Multiple Evidence Domains, Not Just One

There is no consensus about what "evidence" or "evidence-based policy" includes. Evidence comes from a range of systematic analyses and many different types of research, depicted in a simple way by the Venn diagram (exhibit 1): performance measurement, program evaluation, statistical analysis, and policy analysis. Program administrators and policymakers use all types of information to make decisions. Researchers and evaluators should provide the best evidence possible.

2100 M Street NW
Washington DC 20037
urban.org

EXHIBIT 1

Evidence = Evaluation + Performance Measurement + Policy Analysis + Statistical Analysis



Performance measurement. Many federal agencies have adopted performance management systems, at the urging of the Office of Management and Budget (OMB) and encouraged by the Government Performance and Results Act of 1993 and the subsequent GPRA Modernization legislation enacted in 2010. Each agency is at a different stage of development, but in general they follow many of the strategies common in private-sector and government performance, such as PerformanceStat:

- set quantitative targets and goals associated with objectives and mission,
- use performance metrics (e.g., activities, outputs, outcomes) to track progress,
- analyze data to determine achievement of targets,
- hold regular management meetings to review results and determine followup actions needed to improve performance, and
- take action to improve performance.

In short, performance measurement in the federal government is a management function, just as it in the private sector, and is performed within an organization or agency on a regular basis, using data on program activities and results.

Program evaluation. In the federal government, program evaluation and program performance measurement are related, since both assess the results of public programs or policies (thus the overlap

in the Venn diagram). They may both also use program and administrative data. But the research questions, analytic methodologies, and interpretations are different.

Performance measurement requires asking whether a program or agency is reaching its goals and targets, and tracking progress on an ongoing basis. Evaluations may address various questions, including determining whether a program or service strategies has positive impact or "added value."

Formal evaluations are conducted periodically at strategic times, designed to address specific questions using rigorous designs and analytic methods. The most rigorous evaluations are those that estimate the impact of a program or policy compared to what would have happened otherwise (that is, compared to the counterfactual), using experimental randomized controlled trials or high-quality nonexperimental statistical methods. While performance analysis is done by managers and staff inside an organization, evaluations are conducted by independent researchers, usually outside the organization, who have no direct stake in the program or the results. The credibility of the results of an evaluation depend on the independence of the researchers conducting the study.

Statistical analysis and surveys. Statistical analysis and surveys are also critical sources of evidence about broader trends, domains, populations, and contexts. Data for conducting analysis are often used in performance measurement and in formal program evaluations as well as being presented in analytic tables, papers, and reports. The Bureau of Labor Statistics, for example, produces statistical data and reports on economic conditions, including the monthly unemployment rate, and information on average employment and earnings for the nation, regions, states, and substate geographic areas, and for different populations and subgroups. Special surveys, often done in partnership across statistical agencies such as BLS and the Census Bureau, also provide nuanced data on households, individuals, workers, firms, and industries. Analyzing the nation's statistical data provides the foundation for much evidence, and can be merged with other data for formal evaluation or performance purposes.

Policy analysis. Policy analysis examines the interaction among the other types of research and provides comprehensive, usually multidisciplinary, assessment of policy or program alternatives—ideally from every angle possible—to inform decisions. Policy memoranda, literature reviews, background analysis, white papers, legislative reviews, and other analyses help lay the groundwork for more formal impact evaluations and delineate the context within which all findings should be considered. Quantitative and qualitative policy analysis takes into account the existing research findings and perspectives of various interest groups, political parties, and other stakeholders, and considers legal, fiscal, economic, social, and operational implications of policies or potential policies.

Each type of evidence is useful for decisionmaking. Researchers may be professionally drawn to a particular theoretical or methodological approach. Economists and many evaluators, for example, tend to emphasize experimental and causal impact analysis, because these rigorous approaches produce the most credible and statistically precise estimates of impact compared to what would happen without the program or intervention. Public administration professionals are trained to develop and use management and performance metrics. Other types of analysis, when done well, also provide critical information about, for example, qualitative and quantitative analysis of information about how

programs operate and deliver services and how performance improvement occurs. Such operational information often helps interpret the findings from formal impact evaluations and helps understand variations in performance. Regardless of the type of analysis, the key is to make sure the studies are of the highest quality, rigor, and independence so the evidence that emerges is credible and useful.

Abide by Professional Evaluation Principles

Most professional associations and organizations have an accepted set of standards or principles that guide the conduct or work in their respective fields. Researchers and evaluators generally follow academic freedom and scientific integrity principles that are standard in academic and research fields, such as free inquiry and communication, objectivity, transparency, and avoidance or disclosure of conflicts of interest. Researchers that conduct studies with federal sponsorship are expected to follow the standard principles of their disciplines.

Federal evaluation and research offices also have adopted professional standards, and they typically follow the same academic and scientific integrity principles in place in federal agencies. For example, the Department of Labor's formal evaluation policy statement consists of the principles of methodological rigor, independence, transparency, ethics, and relevance (exhibit 2). The policy statement was formally adopted in 2013, drawing from principles that also exist in the scientific, research, and evaluation professions, and consistent with what is taught in graduate courses on program evaluation and with what other federal departments have adopted. The Administration on Children and Families at the Department of Health and Human Services, the Institute for Educational Sciences at the Department of Education, the Bureau of Labor Statistics at the Department of Labor, the Environmental Protection Agency, and the Centers for Disease Control and Prevention have similar policies.

EXHIBIT 2

Key Principles in USDOL's Evaluation Policy Statement

- Rigor—use the most rigorous and appropriate methods possible; technical peer review
- Independence—objectivity in design, conduct, analysis, and release of results free of undue influence
- Transparency—release all reports, results; broad dissemination; public use data files
- Ethics—human subjects protections, security, privacy, professional standards; comply with spirit and letter of law and rules
- Relevance—programs, policymakers, leadership, stakeholders, state/local/tribal partners

Having a formal statement ensures that all administrators and staff know, agree with, and understand the role of evaluation and resulting findings and reports. Research must be objective, methodologies must follow rigorous standards, all studies and results are peer reviewed, and the independent reports are all publicly released. Evaluators must understand the principles in effect in each federal agency, follow them, while also abiding by their own professional evaluation standards and academic scientific integrity principles.

Carefully Determine What Will Be Studied

Perhaps the most essential evaluation principle, and the most difficult to institutionalize in a federal agency, is relevance: how to make sure evaluations are relevant to policymakers, program administrators, and staff, and relevant to the political priorities of the current administration and leadership. If evidence and evaluation activities are not relevant, then the studies and their findings become marginalized in the government—although researchers may still enjoy discussions among academic peers and publish in scholarly journals.

Federal agencies approach the relevance challenge in various ways, particularly by obtaining input from stakeholders in setting an agenda. At the Department of Labor, a strategic participatory Learning Agenda approach is used to decide what will be evaluated (exhibit 3). Each operating agency in the department, in collaboration with the Chief Evaluation Office, updates their learning agenda annually—identifying priority issues, topics, and sometimes particular studies. The evaluation office then takes all the agendas and priorities and the budget available, and prepares a Departmental evaluation plan. The plan is published in the *Federal Register* for input and comment from stakeholders and the public, and Congress and OMB are notified of the planned priority studies (exhibit 4). This participatory approach to deciding what to evaluate helps make sure the studies are of interest to the agencies, program administrators and leaders, policymakers, and stakeholders.

Other federal evaluation offices use expert advisory groups to help develop evaluation priorities (e.g., the Administration on Children and Families), extensive public outreach and requests for information (e.g., Housing and Urban Development's Road Map), or stakeholder and expert outreach (e.g., the Employment and Training Administration's five-year research plan).

Researchers and evaluators should understand the ways federal agencies decide on which studies to sponsor and make sure the resulting studies are as relevant as possible to administrators, policymakers, and other stakeholders—not just that the study meets current academically accepted standards.

EXHIBIT 3

Strategic Participatory Learning Agenda Approach

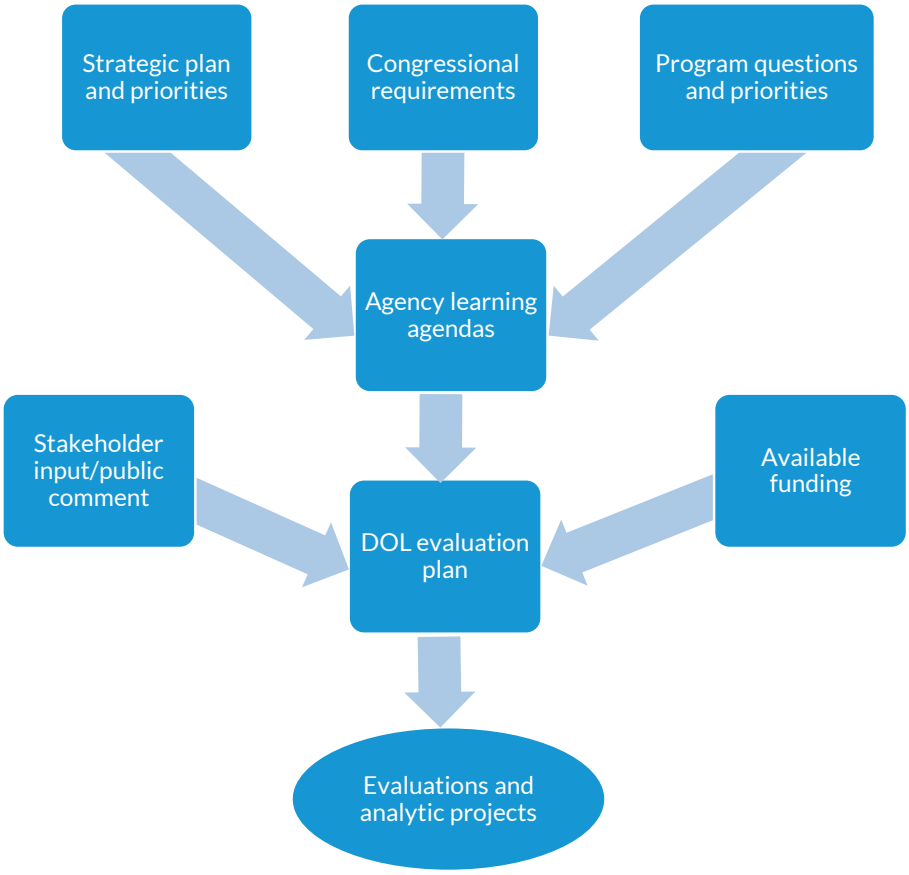


EXHIBIT 4

Key Principles in USDOL's Evaluation Policy Statement

- Literature/background reviews
- Systematic evidence reviews
- Feasibility and exploratory studies; evaluability assessments; concept testing; proof of concept
- Behavioral insights/behavioral economics/ "nudge" tests
- Implementation studies
- Performance measurement analyses
- Statistical and survey analyses
- Formal impact evaluations (experimental and nonexperimental)

Another way evaluators can be relevant is to understand the government's history around evidence, including some skepticism and anxiety about evaluation. Recognize that many in the government are apprehensive about evaluation, performance measures, and evidence in general—even though evidence is now clearly a high priority in the executive and legislative branches. Many program administrators and staff worry that an evaluation's findings may be misinterpreted or taken as an "up or down" vote on their budget or their program. They worry about negative findings. Some worry that evaluation or research results may not support a particular high priority policy, or a sensitive or controversial topic, and could lead to political difficulties; some worry that the evaluators do not really understand their programs. But without good evaluations and research, agencies lack useful information that could help them improve their programs.

There are ways to ease the apprehension. Some agencies have a long history of doing and using evaluations. They know what it takes and understand both the risks and benefits. Others have little experience with formal evaluations. Evaluators should strive to produce useful reports which can help promote an environment where agencies and staff want to use evidence and data. Often that means starting with small steps. For example, implementation evaluations and evaluability assessments are good initial steps towards more quantitative evaluations and the findings help inform programs about possible ways to improve operations and results. Another way to prepare staff for evaluations is for researchers to work collaboratively with program offices and staff to develop logic models or similar conceptual frameworks in designing a study, thus engaging operational staff in the study.

In addition, rather than doing only large-scale, random-assignment experimental impact evaluations, it makes sense to also carry out short-term, smaller-scale evaluations that use rigorous methods. Findings are produced fairly quickly providing staff with real results in a timely manner. Behavioral insights or "nudge" tests, for example, have been well received in federal and state agencies, probably because the findings come quickly, usually focus on how to improve a process or message, and tend not produce a positive or negative impact estimate for the program as a whole. Efficacy tests and planned variations, using randomized controlled trials, also gain more support than programwide evaluations, and the results can be useful in helping administrators make decisions about how to improve program performance and results.

Explain and Help Interpret Findings for Policymakers and Program Administrators

The next and related point is that it is important that evaluators explain and help interpret findings realistically. Regardless of the study or the number of cumulative studies on the same topic, the findings need to be interpreted in a manner that nonevaluators can understand them and use them. The interpretation of evaluation findings is often very difficult for program administrators and policymakers who are appropriately more focused on implementing their programs and policies. Their tendency may be to see results from one study and try to determine how to take action or make decisions based on that one study, or to consider making budgetary decisions based on one study. It is the job of evaluators and policy analysts to do more realistic and thoughtful interpretation of findings from a body of evidence—and to avoid inappropriate emphasis on any single study.

At the Department of Labor, this interpretive role plays out in various ways. When preparing annual agency budget proposals, administrators are encouraged to confer with the Chief Evaluation Office, where evaluation specialists help draft evidence justifications for their budget requests. The Department also has an evidence-based clearinghouse called CLEAR (Clearinghouse for Labor Evaluation and Research) that includes evaluation reviews and ratings of the quality of each study using systematic standards and guidelines. CLEAR also includes short syntheses of accumulated research on topics that are relevant to agencies. In addition, evaluation contractors are routinely asked to prepare policy briefs or practice briefs for nontechnical audiences that highlight lessons and implications for program administrators, staff, and policymakers. These types of activities focus on summarizing and communicating findings for nonevaluators.

Interpreting and communicating results is an ongoing challenge that has to be a priority for evaluation staff within government and for researchers and evaluators doing evaluations—if they want their studies to be relevant, heard in the policy arena, and used properly and not misused.

Fill the Pipeline of Future Evaluators

The final point relates to the importance of having a professional pipeline of evaluators who understand public agencies and public policy, and a pipeline of professional public administrators who understand evaluation. Twenty years ago when the nation first started hearing about evidence-based policy analysis, the number of academically trained policy analysts was only a fraction of what it is today. Public policy and public administration graduate schools do a great job at preparing master's and doctoral students for conducting evaluations of public programs, for understanding and critically reviewing evaluation reports, and preparing some to serve in the government. While many public policy professionals conduct evaluations, most major evaluations are carried out by researchers from other disciplines, primarily economics and sociology. Graduates of public policy and public administration schools are not the only ones who do evaluations, though, and public policy schools could think of ways to partner more directly with other schools and departments that are training economists, statisticians, sociologists, psychologists, and others who often do program evaluations. Public policy faculty can connect public policy relevance to the technical and theoretic training all students receive. Schools of business, management, international affairs, and operations research could consider better ways to apply their essential theories and tools to evaluating public policy issues and topics, for example, by requiring graduate students to take program evaluation classes. The interdisciplinary nature of public policy and public administration schools and programs provides a natural ground for cross-fertilization, and should teach evaluation methods in ways that are relevant to other disciplines. Many federal departments have research grant programs for academics from any discipline, which is one way to build the pipeline of policy-relevant academic researchers. More cross-disciplinary academic coordination could play an important role in building a pipeline of policy-relevant researchers **and** a pipeline of evidence-sensitive public managers.

Demetra Smith Nightingale is an Institute fellow at the Urban Institute, where her research focuses on social, economic, and labor policy issues. She was the chief evaluation officer at the US Department of Labor from 2011 to 2016, where she developed what is recognized as one of the premier evaluation units in the federal government. This paper is based on a presentation Nightingale gave at the awards luncheon for the Association for Public Policy Analysis and Management, November 4, 2016, in Washington, DC.

URBAN INSTITUTE

2100 M Street NW
Washington, DC 20037

www.urban.org

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is dedicated to elevating the debate on social and economic policy. For nearly five decades, Urban scholars have conducted research and offered evidence-based solutions that improve lives and strengthen communities across a rapidly urbanizing world. Their objective research helps expand opportunities for all, reduce hardship among the most vulnerable, and strengthen the effectiveness of the public sector.