Evaluation and Analytics to Improve, Not Just Evaluate, Program Performance

Batia Katz
URBAN INSTITUTE

Shelley Metzenbaum
THE BETTER PROJECT

October 2020

Evidence-based policymaking and program evaluation are quickly gaining traction in the federal government. Some agencies did evaluations and analytics before the 2019 Evidence Act was passed,[1] but now all federal agencies must boost their evidence and evaluation capacity. The law mandates that federal agencies designate an evaluation officer to assure high-quality evaluations, improve public access to government data, and craft a learning agenda identifying and prioritizing knowledge gaps to fill and the means to fill them.

The 2019 Evidence Act's purpose is to build and share knowledge relevant to the objectives federal agencies seek to advance, and then apply that knowledge to federal funding and other activities to realize better government and real-world results.[2] As agencies create learning agendas for themselves, conduct evaluations, and share and analyze their data, they will be able to gather and disseminate important evidence on which programs advance progress and which do not.

But will this evaluation and evidence-building capacity also be used to improve program performance?

Speaking at the Urban Institute on January 23rd, 2020, former Volcker Alliance Founding President and OMB Associate Director for Performance and Personnel Management Shelley Metzenbaum asked this exact question. Expanding on a panel about how agencies can engage external stakeholders in creating their learning agendas,[3] Dr. Metzenbaum suggested that the questions asked in learning

agendas should not only focus on whether or not a program works, but also—and arguably more important—on how to improve the program.

Evidence, including that from data analytics and evaluations, can enhance public programs[4] in many ways. It helps agencies detect problems and opportunities, set priorities, design treatments, assess programs and practices, and adapt processes for continual improvement.

Analyses of performance and other information can help answer a number of questions or, at least, start to pave the path to answers. What causal factors can be influenced, such as reading habits or safety belt use or smoking, to improve outcomes or reduce problems? Can positive outliers be found— the "moneyball" programs that outperform their peers? Can practices used by "moneyball players" be identified that, when replicated, yield similarly positive results in other situations? These practices might be found, for example, in approaches to recruitment and retention, coaching, and behavior change campaigns. Do clusters of similar problems exist that are likely to respond well to cost-saving group interventions? Is money going to where the biggest problems, or biggest opportunities for gain, are? Are providers in different locations all making progress at a reasonable pace and, if they are not, can the reason for those difference be discerned? These are the kinds of questions programs can be asking; then using analyses of performance and other data to answer and inform next steps.

Program evaluation is distinct from but complementary to performance measurement and management. Performance measurement and management are done continuously to improve daily program operations. Did a job-training program succeed in getting trainees in the door? Did employment results vary by trainer or by trainee characteristics? Performance managers ask such questions. They analyze data to look for anomalies, positive outliers, relationships, clusters, and other patterns to figure out what is going on, discern relationships, and decide next steps to fix problems and pursue opportunities. They also use the data to inform priorities.

Evaluation complements performance measurement. It can more rigorously isolate the impact of government and grantee actions from the impact of other causal factors and gauge the size of those impacts. It can also discern the relative import of causal factors. For example, were reductions in youth pregnancy the result of government campaigns, entertainment programs about youth pregnancy, new medical methods, or some or all of the above?

Both performance management and evaluations have great strengths when properly used but must be used appropriately and cautiously—resisting the temptation to conclude that a program or practice did or did not work without taking into consideration when, for what, and for whom it worked and for whom it did not. For example, a regression analysis[5] finding that a 1 percent increase in inspections raised compliance levels 10 percent cannot predict that a ten-fold increase in inspections will boost compliance to 100 percent. Moreover, practice variations as subtle as the timing of an inspection may affect the impact of those practices, as an office of the US Coast Guard discovered when it shifted its inspections from daytime to nighttime after realizing that most of the oil spills it wanted to stop happened at night. Also, care needs to be taken not to draw premature conclusions about programs whose full effects make take a long time to be fully realized. A later evaluation of the Moving to

Opportunity (MTO) program,[6] which relocated families from neighborhoods with high poverty levels to those with lower poverty levels, showed larger and different effects than those found by earlier evaluations. Care must also be taken not to mandate prematurely widescale use of practices that appear effective in one study before demonstrating replicability in another situation, and, if possible, understanding the circumstances where replication is likely to work well and where it is not.

Evaluation tends to be more discrete and episodic than performance measurement, although a growing appreciation for rapid-cycle testing, perhaps influenced by user-centered software design that iteratively tests and assesses different human interface options compared with each other, is starting to change that, more seamlessly integrating the use of and findings from smaller scale well-designed measured trials in performance management routines. Rapid-cycle testing tries new practices on a small scale, rapidly assesses their impact, adjusts actions based on assessment findings, and quickly assesses the effect of the adjusted actions. It helps fine-tune program and practice design but also increases discovery of improvement opportunities that analyses of past performance cannot reveal. It allows program providers to test to find better practices in real time.

Advances in how big data can be integrated and analyzed bring unprecedented opportunities for applying insights from analytics combined with well-designed, measured real-time field trials. A New York City data analytics team, for example, wanted to help inspectors reduce building fires and other safety problems. Using data from 20 city databases plus interviews with field inspectors, they identified factors possibly correlated with unsafe buildings. With this analysis, they proposed a new schedule for buildings inspections, which building inspectors then tried. The prior scheduling method found conditions serious enough to warrant a vacate order 13 percent of the time. The new, data-informed schedule led inspectors to issue vacate orders for 70 percent of inspected buildings. [7]

Performance measurement and management, data analytics, and program evaluation can and should enhance one another, complemented by learning agendas—the focus of the Urban Institute event.[8] Learning agendas establish routines for identifying knowledge gaps to fill, prioritizing their importance, and deciding on the means to fill them, including filling data gaps and strengthening analytic and evaluation capacity.[9] Used together, these tools will improve the performance and impact of government programs. These tools promise to produce especially robust returns when continuous learning and improvement communities are created where they don't already exist and supported where they do to connect program officials in central and regional offices, frontline workers wherever they work, researchers, performance measurement/management and evaluation experts, information technology leaders, and data analysts to work together to learn from experience, collaborate on well-designed trials, and decide next steps.

Notes

1   Foundations for Evidence-Based Policymaking Act of 2018, H. R. 4174, https://www.congress.gov/bill/115th-congress/house-bill/4174; Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115-435, 132 Stat. 5529 (2019), https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf.

2   Jessica Shakesprere, "A New Law Could Help the Government Rely on Evidence to Improve Programs," *Urban Wire* (blog), February 5, 2019, https://www.urban.org/urban-wire/new-law-could-help-government-rely-evidence-improve-programs.

3   "Delivering on the Evidence Act: How Agencies Can Engage Stakeholders in the Learning Agenda Process," Urban Institute (Event), January 23, 2020, https://www.urban.org/events/delivering-evidence-act-how-agencies-can-engage-stakeholders-learning-agenda-process.

4   Demetra Nightingale, "How Evidence-Based Decisionmaking Enhances Public Service," *Urban Wire* (blog), November 27, 2017, https://www.urban.org/urban-wire/how-evidence-based-decisionmaking-enhances-public-service.

5   Regression analysis is a statistical process used to estimate relationships between variables of interest. Program evaluators might use regression analyses to better understand the effects of their programs on a certain outcome.

6   "A Summary Overview of Moving to Opportunity: A Random Assignment Housing Mobility Study in Five U.S. Cities," National Bureau of Economic Research, accessed July 31, 2020, https://www.nber.org/mtopublic/MTO%20Overview%20Summary.pdf.

7   Viktor Mayer-Schönberger and Kenneth Cukier, "Big Data in the Big Apple," *Slate*, March 6, 2013, https://slate.com/technology/2013/03/big-data-excerpt-how-mike-flowers-revolutionized-new-yorks-building-inspections.html.

8   "Delivering on the Evidence Act: How Agencies Can Engage Stakeholders in the Learning Agenda Process," Urban Institute (Event).

9   Kriti Ramakrishnan and Keith Fudge, "Four Things to Know About Learning Agendas," *Urban Wire* (blog), November 7, 2018, https://www.urban.org/urban-wire/four-things-know-about-learning-agendas.

About the Authors

Batia Katz is a research assistant in the Center on Labor, Human Services, and Population at the Urban Institute, where she researches workforce development. Katz's previous research experience includes studying the science labor market, the impact of personality traits on employment outcomes, and gender and family in the workforce. Katz graduated with high honors from Haverford College, where she earned a BA in economics.

Shelley Metzenbaum helps governments and others to use and communicate goals, data, analyses, and research to improve outcomes, and to avoid use of these tools in ways that create fear and dysfunctional responses. Metzenbaum headed the Office of Performance and Personnel Management at the US Office of Management and Budget during the first term of the Obama Administration, and subsequently served as founding president of The Volcker Alliance. Current projects include advising the National Head Start Association; multiple efforts to improve federal grant outcomes, operational quality, and transparency; regulatory metrics; and advising an effort to transform correctional oversight.

Acknowledgments

URBAN INSTITUTE

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.