

# BACKGROUND

No. 3063 | OCTOBER 15, 2015

## Evidence-Based Policymaking: A Primer

David B. Muhlhausen, PhD

### Abstract

*Every American should be concerned about the nation's extraordinary level of debt. Congress, which in recent years has seemed incapable of curbing spending and allocating resources effectively, needs to relearn how to be a wise steward of the federal purse. One obvious tool missing from the budget-cutters' toolbox is evidence-based policymaking. When practiced correctly, evidence-based policymaking is a tool that would allow policymakers to base funding decisions on scientifically rigorous impact evaluations of programs. Given scarce federal resources, federal policymakers should fund only those programs that have been proven to work and defund programs that do not work.*

The federal government's total debt exceeds \$18.1 trillion.<sup>1</sup> However, even this huge figure understates the nation's debt. Due to the absence of fiscal restraint after restoration of the debt limit in March 2015 at a level of \$18.1 trillion, the Department of the Treasury has been forced to engage in extraordinary measures to continue borrowing, mainly raiding funds required to pay future federal employee retirement obligations.<sup>2</sup> Needless to say, such measures cannot continue forever. Cuts in federal spending are essential to restore long-term balance to federal accounts.

One obvious tool missing from the budget-cutters' toolbox is a means of accurately evaluating the effectiveness of various government programs. The sad truth is that federal social programs in particular are almost never held accountable for their performance, instead being justified primarily by their good intentions. The political process of deciding public policy should be informed not only by

### KEY POINTS

- The effectiveness of federal programs is often unknown. Many programs operate for decades without ever undergoing thorough scientific evaluations.
- To break the culture of wasteful spending, Members of Congress need to relearn how to be wise stewards of the federal purse.
- One obvious tool missing from the budget-cutters' toolbox is evidence-based policymaking.
- When practiced correctly, evidence-based policymaking is a tool that would allow policymakers to base funding decisions on scientifically rigorous impact evaluations of programs.
- Changing the federal government's emphasis on measuring success by the amount of spending and good intentions will not be easy. However, enacting real evidence-based reforms would be a step in the right direction in changing the culture in Washington toward funding programs that work and defunding those that do not work.

This paper, in its entirety, can be found at <http://report.heritage.org/bg3063>

**The Heritage Foundation**  
214 Massachusetts Avenue, NE  
Washington, DC 20002  
(202) 546-4400 | [heritage.org](http://heritage.org)

Nothing written here is to be construed as necessarily reflecting the views of The Heritage Foundation or as an attempt to aid or hinder the passage of any bill before Congress.

values, but also by rigorous evidence. Caring about a particular social problem and, in many cases, spending billions of dollars on it will not necessarily alleviate the social problem.

In fact, the effectiveness of federal programs is often unknown. Many programs operate for decades without ever undergoing thorough scientific evaluations.

Given scarce federal resources, federal policymakers need to fund programs that work and defund programs that do not work. Americans, who fund these programs with hard-earned tax dollars, deserve better than Congress's current habit of funding programs that may not produce their intended results.

To plug this information gap, the evidence-based policy movement seeks to inform policymakers through scientifically rigorous evaluations of the effectiveness of government programs.<sup>3</sup> In other words, the movement provides tools to identify what works and what does not work.

### What Is Evidence-Based Policymaking?

Evidence-based policymaking would base funding decisions on scientifically rigorous impact evaluations of programs. Rigorous impact evaluations that use randomized controlled trials provide policymakers with improved capability to oversee government programs and be more effective stewards of the federal purse. There is little merit in continuing programs that fail to ameliorate their targeted social problems. Programs lacking evidence of effectiveness or those that do not work at all do not deserve continued funding. Congress needs to take the lead in ensuring that the social programs that it funds are evaluated.

*First*, when authorizing a new social program or reauthorizing an existing program, Congress should specifically mandate multisite experimental

evaluations of the program.<sup>4</sup> Experimental evaluations are the only way to determine with a high degree of certainty the effectiveness of social programs. Thus, Congress should mandate that all recipients of federal funding, if selected for participation, must cooperate with evaluations in order to receive future funding.

### When authorizing a new social program or reauthorizing an existing program, Congress should specifically mandate multisite experimental evaluations of the program.

*Second*, the experimental evaluations should be large-scale, nationally representative, multisite studies. When Congress creates social programs, the funded activities are intended to be spread across the nation. For this reason, Congress should require nationally representative, multisite experimental evaluations of these programs. For multisite evaluations, the selection of the sites to be evaluated should be representative of the population of interest for the program. When program sites and sample participants are randomly selected, the resulting evaluation findings will have a high degree of validity.

Large-scale experimental evaluations based on multiple sites avoid problems of simplistic generalizations. A multitude of confounding factors that vary by location can influence the performance of social programs.<sup>5</sup> What works in Tulsa, Oklahoma, may not work in Baltimore, Maryland. Thus, the larger the size of the evaluation (e.g., sample size and number of sites), the more likely the federal social program will be assessed under all of the conditions under which it operates.

1. U.S. Department of the Treasury, "The Debt to the Penny and Who Holds It," <http://treasurydirect.gov/NP/debt/current> (accessed March 9, 2015).
2. Romina Boccia and Michael Sargent, "The Debt Limit Is Back—What Now?" *The Hill*, March 17, 2015, <http://thehill.com/blogs/pundits-blog/economy-budget/235875-the-debt-limit-is-back-what-now> (accessed July 21, 2015), and D. Andrew Austin, "The Debt Limit: History and Recent Increases," Congressional Research Service *CRS Report*, July 1, 2015.
3. See Karen Bogenschneider and Thomas J. Corbett, *Evidence-Based Policymaking: Insights from Policy-Minded Researchers and Research-Minded Policymakers* (New York: Routledge, 2010).
4. For a detailed discussion of the merits of multisite experimental evaluations, see David B. Muhlhausen, "Evaluating Federal Social Programs: Finding Out What Works and What Does Not," Heritage Foundation *Backgrounder* No. 2578, July 18, 2011, <http://www.heritage.org/research/reports/2011/07/evaluating-federal-social-programs-finding-out-what-works-and-what-does-not>.
5. David B. Muhlhausen, *Do Federal Social Programs Work?* (Santa Barbara, CA: Praeger, 2013).

While individual social programs operating in a single location and funded by the federal government may undergo experimental evaluations, these small-scale, single-site evaluations do not provide good information on the general effectiveness of national social programs. Small-scale evaluations assess only the impact on a small fraction of the people served by federal social programs. The success of a single program that serves a particular jurisdiction or population does not necessarily mean that the same program will achieve similar success in other jurisdictions or among different populations. Simply, small-scale evaluations are poor substitutes for large-scale, multisite evaluations.

Thus, federal social programs should be evaluated in multiple sites to test them in the various conditions in which they operate and in the numerous types of populations that they serve. In addition, a multisite experimental evaluation that examines the performance of a particular program in numerous and diverse settings can potentially produce results that are more persuasive to policymakers than results from a single locality.<sup>6</sup>

### Why Random Assignment Is Crucial

Determining the impact of social programs requires comparing the outcomes for those who received assistance with the outcomes for an equivalent group that did not experience the intervention. However, evaluations differ by the quality of methodology used to separate the net impact of programs from other factors that may explain differences in outcomes between those receiving treatment and those not receiving treatment.<sup>7</sup>

Experimental evaluations that use random assignment are the “gold standard” of evaluation designs. Randomized experiments attempt to demonstrate causality by (1) holding all other possible causes of the outcome constant; (2) deliberately altering only the possible cause of interest; and (3) observing whether the outcome differs between the intervention and control groups.

When conducting an impact evaluation of a social program, identifying and controlling for all the possible factors that influence the outcomes of interest is generally impossible. We simply do not have enough knowledge. Even if we could identify all possible causal factors, collecting complete and reliable data on all of these factors would likely still be beyond our abilities. For example, it is impossible to isolate a person participating in a social program from his family in order to “remove” the influences of family. This is where the benefits of random assignment become clear.

Because we do not know enough about all possible causal factors to identify and hold them constant, randomly assigning test subjects to intervention and control groups allows us to have a high degree of confidence that these unidentified factors will not confound our estimate of the intervention’s impact. Random assignments should evenly distribute these unidentified factors between the intervention group and the control group of an experimental evaluation.

However, the benefits of random assignment are most likely to occur with large sample sizes. Randomized evaluations using small sample sizes do not have the same scientific rigor as randomized evaluations using large sample sizes. Random assignment helps to ensure that the control group is equivalent to the intervention group in composition, predispositions, and experiences. The groups are composed of the same types of individuals in terms of their program-related and outcome-related characteristics. In addition, members of both groups should be similarly disposed toward the program. Further, the intervention and control groups should have the same experiences regarding time-related variables, such as their maturity level and history.<sup>8</sup>

Randomized experiments have the highest internal validity when sample sizes are large enough to ensure that idiosyncrasies that can affect outcomes are evenly distributed between the program and control groups. With small sample sizes, disparities

---

6. Erica B. Baum, “When the Witch Doctors Agree: The Family Support Act and Social Science Research,” *Journal of Policy Analysis and Management*, Vol. 10, No. 4 (Autumn 1991), pp. 603–615, and Judith M. Gueron, “The Politics of Random Assignment: Implementing Studies and Affecting Policy,” in Frederick Mosteller and Robert Boruch, eds., *Evidence Matters: Randomized Trials in Education Research* (Washington, DC: Brookings Institution, 2002), pp. 15–49.

7. For a detailed discussion of evaluation methodology, see Muhlhausen, *Do Federal Social Programs Work?*

8. The internal validity threat of history occurs when events taking place concurrently with the intervention could cause the observed effect, while maturation occurs when natural changes in participants that occur over time could be confused with an observed outcome. For a more detailed discussion of threats to internal validity, see Muhlhausen, *Do Federal Social Programs Work?*

in the program and control groups can influence the findings. For this reason, evaluations with large samples are more likely to yield scientifically valid impact estimates.

### Evidence-Based Policymaking in Practice

In practice, policymakers frequently assume when something has been found effective in one setting, the same results can be repeated elsewhere. However, the history of social programs is replete with examples of programs effective in one location that simply did not work elsewhere.

The federal government has a poor record of replicating effective social programs.<sup>9</sup> An excellent example of a federal attempt to replicate an effective local program is the Center for Employment Training Replication.<sup>10</sup> Of 13 youth job-training programs evaluated, the JOBSTART demonstration found only one program to have a positive impact on earnings: the Center for Employment Training (CET) in San Jose, California.<sup>11</sup> Based on the results for the CET, the U.S. Department of Labor replicated and evaluated the impact of CET in 12 other sites using random assignment.<sup>12</sup> The CET model had little to no effect on short-term and long-term employment and earnings outcomes at these other locations. According to the evaluation's authors, "[E]ven in sites that best implemented the model, CET had no overall employment and earnings effects for youth in the program, even though it increased participants' hours of training and receipt of credentials."<sup>13</sup>

Just because an innovative program appears to have worked in one location does not mean that the program can be effectively implemented on a larger scale. Proponents of evidence-based policymaking should not automatically assume that allocating

taxpayer dollars toward programs attempting to replicate previously successful findings will yield the same results.

The faulty reasoning that drives such failed expansions of social programs is known as the "single-instance fallacy."<sup>14</sup> This fallacy occurs when a person believes that a small-scale social program that appears to work in one instance will yield the same results when replicated elsewhere. Compounding the effects of this fallacy, we often do not truly know why an apparently effective program worked in the first place. In particular, the dedication and entrepreneurial enthusiasm of a program's founder may be difficult to quantify or duplicate. Thus, how can we know whether we can replicate success elsewhere?

---

### Just because an innovative program appears to have worked in one location, does not mean that the program can be effectively implemented on a larger scale.

---

One way is to deem program models as "evidence-based" only after they have been found by experimental evaluations to have consistent statistically significant effects that meaningfully ameliorate the targeted social problem in at least three different settings. Once a program model has been found to produce meaningful results in multiple settings, the likelihood of its successful replication elsewhere should increase significantly.

Identifying real evidence-based programs is more easily said than done. In the United States Code, the phrase "evidence-based" appears 122 times, frequently without clear meaning.<sup>15</sup> In practice, what

---

9. Muhlhausen, *Do Federal Social Programs Work?*

10. Cynthia Miller et al., *The Challenge of Replicating Success in a Changing World: Final Report on the Center for Employment Training Replication Sites*, Manpower Demonstration Research Corporation, September 2005, <http://www.mdr.org/publication/challenge-repeating-success-changing-world> (accessed September 2, 2015).

11. George Cave et al., *JOBSTART: Final Report on a Program for School Dropouts*, Manpower Demonstration Research Corporation, October 1993, <http://www.mdr.org/project/jobstart> (accessed September 2, 2015).

12. Miller et al., *The Challenge of Replicating Success in a Changing World*.

13. *Ibid.*, p. xi.

14. Stuart M. Butler and David B. Muhlhausen, "Can Government Replicate Success?" *National Affairs*, Spring 2014, pp. 25–39, <http://www.nationalaffairs.com/publications/detail/can-government-replicate-success> (accessed August 4, 2015).

15. Search on U.S. House of Representatives, Office of the Law Revision Counsel, "United States Code," <http://uscode.house.gov/> (accessed August 4, 2015).

## What Worked in Minneapolis Backfired in Omaha, Milwaukee, and Charlotte

The case of police departments performing mandatory arrests in domestic violence incidents is a poignant example of the need to exercise caution when generalizing findings from a single evaluation of a program in one setting. During the 1980s, criminologists Lawrence W. Sherman and Richard A. Berk analyzed the impact of mandatory arrests for domestic violence incidents on future domestic violence incidents in Minneapolis, Minnesota.\* Compared with less severe police responses, the Minneapolis experiment found that mandatory arrests lead to significantly lower rates of domestic violence. Sherman and Berk urged caution, but police departments across the nation adopted the mandatory arrest policy based on the results of one evaluation conducted in one city.

However, what worked in Minneapolis did not always work in other locations. Experiments conducted by Sherman and others in Omaha, Nebraska; Milwaukee, Wisconsin; Charlotte, North Carolina; Colorado Springs, Colorado; and Dade County, Florida, found mixed results.\*\* Experiments in Omaha, Milwaukee, and Charlotte found that mandatory arrests led to long-term *increases* in domestic violence. Apparently, knowing that they would automatically be arrested prompted repeat offenders to become more abusive. It seems that the following sick logic occurred: If the offender is going to automatically spend the night in jail, then he might as well beat his wife or girlfriend extra good.

In a subsequent analysis of the disparate findings, Sherman postulated that arrested individuals who lacked a stake in conformity within their communities were significantly more likely to engage in domestic violence after arrest, while married and employed arrested individuals were significantly less likely to commit further domestic violence infractions.\*\*\* Thus, mandatory arrest policies may work better in communities with higher rates of marriage and employment, than in communities with lower rates of marriage and employment.

\* Lawrence W. Sherman and Richard A. Berk, "The Specific Deterrent Effects of Arrest for Domestic Assault," *American Sociological Review*, Vol. 49, No. 2 (April 1984), pp. 261-272.

\*\* Lawrence W. Sherman, *Domestic Violence: Experiments and Dilemmas* (New York: Free Press, 1992); Lawrence W. Sherman et al., "Crime, Punishment, and Stake in Conformity: Legal and Informal Control of Domestic Violence," *American Sociological Review*, Vol. 57, No. 5 (October 1992), pp. 680-690; and Lawrence W. Sherman et al., "The Variable Effects of Arrest on Criminal Careers: The Milwaukee Domestic Violence Experiment," *Journal of Criminal Law and Criminology*, Vol. 83, No. 1 (Spring 1992), pp. 137-169, <http://scholarlycommons.law.northwestern.edu/jclc/vol83/iss1/5/> (accessed September 2, 2015).

\*\*\* Sherman, *Domestic Violence*.

counts as evidence-based has varying degrees of meaning. At worst, it is a vapid term.

In the worst case, any program that has undergone an evaluation finding "statistically significant" results, no matter how weak the evaluation design or results, may be considered to be evidence-based. For example, the Department of Health and Human Services Maternal, Infant, and Early Childhood Home Visiting (MIECHV) program—a home-visiting

grants program that provides services to pregnant women and families with children under the age of five—classifies some home-visiting models as "evidence-based," even if the statistically significant effects are few in number, small in magnitude, and have little policy significance.<sup>16</sup>

In other cases, a social program may be categorized as evidence-based based on problematic quasi-experimental studies, even in cases in which

16. Jon Baron, statement before the Subcommittee on Human Resources, Committee on Ways and Means, U.S. House of Representatives, April 2, 2014, [http://waysandmeans.house.gov/UploadedFiles/Jon\\_Baron\\_Testimony\\_HR040214.pdf](http://waysandmeans.house.gov/UploadedFiles/Jon_Baron_Testimony_HR040214.pdf) (accessed August 3, 2015).

experimental evaluations have been undertaken and show, at best, mixed evidence of effectiveness. One particular example is drug courts: a diversionary program for substance-abusing defendants that integrates intensive judicial supervision with mandatory drug testing, incremental sanctions, and substance abuse treatment. A representative of the National Association of Drug Court Professionals labeled drug courts as an “evidence-based” program with effective results “proven beyond a reasonable doubt.”<sup>17</sup> Before such injudicious conclusions are drawn, two key points about the research on the effectiveness of drug courts should be made clear.

---

## The vast majority of drug court evaluations use quasi-experimental designs that suffer from selection bias problems that seriously undercut the scientific validity of the findings.

---

*First*, the vast majority of drug court evaluations use quasi-experimental designs that suffer from selection bias problems that seriously undercut the scientific validity of the findings.<sup>18</sup> For example, many of these quasi-experiments make the questionable comparison of drug court graduates with nongraduates.<sup>19</sup> The “much-heralded findings” based on this faulty methodology “show that the successes succeed and the failures fail.”<sup>20</sup> In other studies, volunteers

are compared with non-volunteers. In either case, quasi-experiments are generally insufficient to the task of separating the effect of drug courts from the individuals’ motivation in succeeding at the outcomes of interest.

*Second*, the experimental evaluations of drug courts do not find consistent effects in reducing recidivism. For example, the experimental evaluation of a Baltimore drug court found that participants had lower one-year and two-year re-arrest rates, but the same participants did not have lower re-conviction or new conviction rates over the same time intervals.<sup>21</sup> In the third year of follow-up, the drug court failed to affect re-arrest rates.<sup>22</sup> In contrast, an experimental evaluation of a drug court in Wilmington, Delaware, “did not find post-treatment differences in outcomes for misdemeanor drug court clients who were assigned to higher versus lower doses of judicial status hearings.”<sup>23</sup> To date, there have not been enough experimental evaluations of drug courts with consistent results to conclude that the approach is effective and replicable.

A lesson to be learned is that policymakers need to be savvy consumers of evaluation research, especially since literature reviews by program advocates may pay little regard to the scientific rigor of the evaluations that they cite. In addition, policymakers need to ensure that programs are evaluated with regard to their impact on the primary purpose for which they have been established. For example, when assessing the impact of prisoner re-entry programs, the most important outcome measure is recidivism. Some

- 
17. Douglas B. Marlowe, testimony before the Subcommittee on Crime and Terrorism, Committee on the Judiciary, U.S. Senate, July 19, 2011, <http://www.judiciary.senate.gov/download/testimony-of-marlowepdf> (accessed July 31, 2015).
  18. David B. Muhlhausen, “Drug and Veterans Treatment Courts: Budget Restraint and More Evaluations of Effectiveness Needed,” testimony before the Subcommittee on Crime and Terrorism, Committee on the Judiciary, U.S. Senate, July 19, 2011, <http://www.heritage.org/research/testimony/2011/07/drug-and-veterans-treatment-courts-budget-restraint-and-more-evaluations-of-effectiveness-needed>.
  19. John S. Goldkamp, Michael D. White, and Jennifer B. Robinson, “Do Drug Courts Work? Getting Inside the Drug Court Black Box,” *Journal of Drug Issues*, Vol. 31, No. 1 (January 2001), pp. 27–72.
  20. *Ibid.*, p. 32.
  21. Denise C. Gottfredson and M. Lyn Exum, “The Baltimore City Drug Treatment Court: One-Year Results from a Randomized Study,” *Journal of Research in Crime and Delinquency*, Vol. 39, No. 3 (August 2002), pp. 337–356; Denise C. Gottfredson, Stacy S. Najjaka, and Brooke Kearley, “Effectiveness of Drug Treatment Courts: Evidence from a Randomized Trial,” *Criminology and Public Policy*, Vol. 2, No. 2 (March 2003), pp. 171–196; Duren Banks and Denise C. Gottfredson, “Participation in Drug Treatment Court and Time to Rearrest,” *Justice Quarterly*, Vol. 21, No. 3 (2004), pp. 637–658; and Denise C. Gottfredson et al., “Long-Term Effects of Participation in the Baltimore City Drug Treatment Court: Results from an Experimental Study,” *Journal of Experimental Criminology*, Vol. 2 (Spring 2006), pp. 67–98.
  22. Gottfredson et al., “Long-Term Effects of Participation in the Baltimore City Drug Treatment Court.”
  23. Douglas B. Marlowe et al., “Are Judicial Status Hearings a ‘Key Component’ of Drug Court? Six and Twelve Month Outcomes,” *Drug and Alcohol Dependence*, Vol. 79, No. 2 (March 2005), p. 154.
-

have questioned the emphasis on recidivism as a measure of effectiveness compared with other measures that assess adjustment or reintegration of former prisoners into society.<sup>24</sup> While intermediate measures, such as finding employment and housing, may be important, these outcomes are not the ultimate goal of re-entry programs. If former prisoners continue to commit crimes after going through re-entry programs, then the programs cannot be judged effective.

### What Congress Should Do

Although every Member of Congress should be concerned about the nation's extraordinary level of debt,<sup>25</sup> the institution in recent years has seemed incapable of curbing spending and allocating resources effectively. To break the culture of wasteful spending, Members of Congress need to relearn how to be wise stewards of the federal purse. There was once a time when Congress placed greater value on fiscal restraint. During the 1950s and early 1960s, the House of Representatives Committee on Appropriations viewed requests for spending by the executive branch as automatically needing to be pared back.<sup>26</sup> As one appropriations subcommittee chairman of that era explained:

When you have sat on the Committee, you see that these bureaus are always asking for more money—always up, never down. They want to build up their organization. You reach a point—I have—where it sickens you, where you rebel against it. Year after year, they want more money.<sup>27</sup>

Rehabilitating a political culture of fiscal restraint will not be easy nor will it be likely accomplished in one fell swoop.

A good first step would be for Congress to identify and fund only those programs that have real evidence of working, while defunding programs

that have been found to fail or that lack evidence of success. A couple of tasks would help Congress to restore fiscal sanity to our nation.

*First*, Congress needs to clearly define “evidence-based.” The term “evidence-based” should mean that experimental evaluations of a program model have found consistent statistically significant effects that meaningfully ameliorate a targeted social problem in at least three different settings.

*Second*, Congress should dramatically increase the number of federal social programs subject to rigorous multisite experimental evaluations to determine what works and what does not work. Adoption of evidence-based policymaking is an important step in helping Congress become wise stewards of the federal purse. To assist in accomplishing this goal, the House of Representatives recently passed the Evidence-Based Policymaking Commission Act of 2015 (H.R. 1831). A companion bill (S. 991) in the Senate has been reported out of the Committee on Homeland Security and Governmental Affairs and is waiting for floor consideration. The legislation's main objective is to advance evidence-based policymaking within the federal government.<sup>28</sup>

*Third*, Congress needs to adopt “pay-for-performance” funding streams. With a pay-for-performance model, current or potential recipients of taxpayer dollars would need to demonstrate through true experimental evaluations that their programs are worthy of continued support. For example, a recipient of a federal delinquency prevention grant would continue to receive future grant funding only if they could demonstrate that their program reduces juvenile delinquency. Currently, the federal government endlessly awards hundreds of billions of dollars in intergovernmental grants each year without requiring grantees to rigorously demonstrate that they are actually solving social problems. A real pay-for-performance model would transform how the federal government works.

24. Christy A. Visher and Jeremy Travis, “Transitions from Prison to Community: Understanding Individual Pathways,” *Annual Review of Sociology*, Vol. 29 (2003), pp. 89–113.

25. U.S. Department of the Treasury, “The Debt to the Penny and Who Holds It.”

26. Richard F. Fenno Jr., *The Power of the Purse: Appropriations Politics in Congress* (Boston, MA: Little, Brown and Company, 1966).

27. *Ibid.*, p. 212.

28. David B. Muhlhausen, “A Commission on Evidence-Based Policymaking: A Step in the Right Direction,” Heritage Foundation *Issue Brief* No. 4363, March 9, 2015, <http://www.heritage.org/research/reports/2015/03/a-commission-on-evidence-based-policymaking-a-step-in-the-right-direction>.

## **Conclusion**

Changing the federal government's emphasis on measuring success by the amount of spending and good intentions will not be easy. However, enacting real evidence-based reforms would be a step in the right direction in changing the culture in Washington toward funding programs that work and defunding those that do not work.

*—David B. Muhlhausen, PhD, is a Research Fellow for Empirical Policy Analysis in the Center for Data Analysis, of the Institute for Economic Freedom and Opportunity, at The Heritage Foundation.*