Congressional Research Service
Informing the legislative debate since 1914

Does Foreign Aid Work? Efforts to Evaluate U.S. Foreign Assistance

Updated July 22, 2016

Congressional Research Service
https://crsreports.congress.gov
R42827

CRS REPORT
Prepared for Members and
Committees of Congress

Summary

In most cases, the success or failure of U.S. foreign aid programs is not entirely clear, in part because historically, most aid programs have not been evaluated for the purpose of determining their actual impact. Many programs are not even evaluated on basic performance. The purpose and methodologies of foreign aid evaluation have varied over the decades, responding to political and fiscal circumstances. Aid evaluation practices and policies have variously focused on meeting program management needs, building institutional learning, accounting for resources, informing policymakers, and building local oversight and project design capacity. Challenges to meaningful aid evaluation have varied as well, but several are recurring. Persistent challenges to effective evaluation include unclear aid objectives, funding and personnel constraints, emphasis on accountability for funds, methodological challenges, compressed timelines, country ownership and donor coordination commitments, security, and agency and personnel incentives. As a result of these challenges, aid agencies do not undertake evaluation of all foreign aid activities, and evaluations, when carried out, may differ considerably in quality.

The Obama Administration has taken several steps to enhance foreign assistance evaluation.

- 2010 Quadrennial Diplomacy and Development Review (QDDR) resulted in, among other things, a stated commitment to plan foreign aid budgets "based not on dollars spent, but on outcomes achieved."
- USAID introduced a new evaluation policy in January 2011.
- The State Department, which began to manage a growing portion of foreign assistance in the 21st century, introduced a new evaluation policy in February 2012, which was updated in January 2015.
- The Millennium Challenge Corporation revised its evaluation policy in 2012, and soon after began releasing its first evaluation reports.

The agency evaluation policies differ in several respects, including their support for impact evaluation, but reflect a common emphasis on evaluation planning as a part of initial program design, transparency and accessibility of evaluation findings, and the application of data to inform future project design and policy decisions. Aspects of the three evaluation policies are compared in the **Appendix**.

Recent reports and policy reviews suggest that aid evaluation frequency and quality have improved in recent years, though progress has been uneven. Attention to this issue remains strong, both within the Administration and among Members of Congress. The 2015 QDDR reemphasizes the role of evaluation, calling for more evaluation training, more strategic use of data, and more timely analysis of lessons learned, among other things. Though recent evaluation reform efforts have been agency-driven, Congress has considerable influence over their impact. Legislators may mandate a particular approach to evaluation directly through legislation (e.g., the Foreign Aid Transparency and Accountability Act, P.L. 114-191, enacted in July 2016), or may support or fail to support Administration policies by controlling the appropriations necessary to implement the policies. Furthermore, Congress will largely determine how, or if, any actionable information resulting from the new approach to evaluations will influence the nation's foreign assistance policy priorities.

Contents

Introduction 1
Why Evaluation? 2
Impact and Performance Evaluations 4
History of U.S. Foreign Assistance Evaluation 5
Evaluation Challenges 11
Applying Evaluation Findings to Policy 17
Current Agency Evaluation Policies 19
Issues for Congress 21
Conclusion 22

Appendixes

Appendix. Select Aspects of Current USAID, State Department, and MCC Evaluation Policies 24

Contacts

Author Information 25

Introduction

In considering budget issues, Congress has long been interested in the relative efficiency and effectiveness of federal programs, including foreign assistance. Foreign assistance evaluation is one aspect of a government-wide effort to link program effectiveness to budgeting decisions. It is also an element of broader foreign aid reforms implemented in recent years. The 2010 Quadrennial Diplomacy and Development Review (QDDR), the basis of many aid policy initiatives, called for the State Department and the U.S. Agency for International Development (USAID) to plan foreign aid budgets and programs "based not on dollars spent, but on outcomes achieved," and for USAID to become "the world leader in monitoring and evaluation."[1] The 2015 QDDR continued the emphasis on evaluation, emphasizing the strategic use of data and the need to build agency evaluation capacity.[2] Rigorous evaluation is also a cornerstone of the Millennium Challenge Corporation (MCC), established in 2004 to promote a new model of development assistance.[3] According to former USAID Administrator Rajiv Shah, global development policies and practices are experiencing a "transformation based on absolute demand for results."[4] That demand comes, in part, from some Members of Congress as they scrutinize the Administration's international affairs budget request and consider foreign aid spending priorities.[5] It also comes from aid beneficiaries and American taxpayers who want to know what impact, if any, foreign aid dollars are having and whether foreign aid programs are achieving their intended objectives.

The current emphasis on evaluation is not new. The importance, purpose and methodologies of foreign aid evaluation have varied over the decades since USAID was established in 1961, responding to political and fiscal circumstances, as well as evolving development theories. There are a number of reasons that this issue has again gained prominence in recent years. For one, foreign aid funding levels increased significantly in the first decade of the 21st century, while evaluations decreased, raising questions about the knowledge basis for aid policy.[6] Analysts have noted that after decades of aid agencies spending billions of dollars on assistance programs, very little is known about the impact of these programs.[7] Some wonder how policymakers can develop effective foreign aid strategies without a clear understanding of how and why prior assistance has succeeded or failed.

This report focuses primarily on U.S. bilateral assistance, not on the work of multilateral aid entities, such as the World Bank, to which the United States contributes. While a wide range of

[1] U.S. Department of State, Quadrennial Diplomacy and Development Review, 2010, *Leading Through Civilian Power*, p. 103.

[2] *Enduring Leadership in a Dynamic World*, the Quadrennial Diplomacy and Development Review, 2015, p. 13.

[3] For more information about the MCC model, see CRS Report RL32427, *Millennium Challenge Corporation*, by Curt Tarnoff.

[4] Statement of USAID Administrator Rajiv Shah to The Cable, as reported in The Cable, June 13, 2012.

[5] While not often discussing evaluation policy per se, some Members appear to be influenced in their policy decisions by their sense of what aid is working and what is not. For example, when introducing her subcommittee's FY2013 proposal at full-committee mark-up on May 17, 2012, House State-Foreign Operations Appropriations Subcommittee Chairwoman Kay Granger remarked that the legislation "only supports programs that work." Senator Lindsay Graham of the Senate State-Foreign Operations Appropriations Subcommittee, explaining the sharp reduction in aid for Iraq in the Senate's FY2013 proposal at a May 22, 2012, mark-up, said "there's no point in throwing good money after bad."

[6] For historic information on foreign aid spending, see CRS Report R40213, *Foreign Aid: An Introduction to U.S. Programs and Policy*, by Curt Tarnoff and Marian L. Lawson.

[7] *When Will We Ever Learn?: Improving Lives Through Impact Evaluation*, Report of the Evaluation Gap Working Group, Center for Global Development, May 2006, p. 1.

federal agencies provide foreign assistance in some form,[8] this report focuses on the three agencies that have primary policy authority and implementation responsibility for U.S. foreign assistance—USAID, the State Department, and the Millennium Challenge Corporation (MCC). It discusses past efforts to improve aid evaluation, as well as ongoing issues that make evaluation challenging in the foreign assistance context. The report also provides an overview of the current evaluation policies of the primary implementing agencies, and discusses related issues for Congress, including recent legislation.

Program Evaluation Government-Wide

Program evaluation is an important issue throughout the U.S. government, and foreign assistance evaluation is just one part of a broader effort by the federal government to improve accountability and program performance through stronger evaluation processes. With the Government Performance and Results Act (GPRA) of 1993, Congress established unprecedented statutory requirements regarding the establishment of goals, performance measurement indicators, and submission of related plans and reports to Congress for its potential use in policy development and program oversight. The GPRA Modernization Act of 2010 updated the original law, requiring more frequent plan updates and on-line posting of data.[9] State Department and USAID strategic planning and assessment documents required by GPRA are available at Performance.gov. The agency-specific evaluation plans discussed in this report are intended to comply with and build upon this government-wide effort.

Why Evaluation?

To know whether aid is successful, one must understand its purpose. The Foreign Assistance Act (FAA) of 1961 (P.L.87-195), as amended, is the authorizing legislation for most modern foreign aid programs. The FAA declared that

> the principal objective of the foreign policy of the United States is the encouragement and sustained support of the people of developing countries in their efforts to acquire the knowledge and resources essential to development, and to build the economic, political, and social institutions that will improve the quality of their lives.[10]

The original legislation lists five principal goals for foreign aid: (1) the alleviation of the worst physical manifestations of poverty among the world's poor majority; (2) the promotion of conditions enabling developing countries to achieve self-sustaining economic growth and equitable distribution of benefits; (3) the encouragement of development processes in which individual civil and economic rights are respected and enhanced; (4) the integration of the developing countries into an open and equitable international economic system; and (5) the promotion of good governance through combating corruption and improving transparency and accountability.[11] Amending legislation over the years added dozens of new, though often overlapping, aid objectives. For example, "the suppression of the illicit manufacturing of and trafficking in narcotic and psychotropic drugs" was added in 1971,[12] "to alleviate human suffering caused by natural and manmade disasters" was added in 1975,[13] and "to enhance the antiterrorism skills of friendly countries by providing training and equipment" and "to strengthen the bilateral

[8] According to ForeignAssistance.gov, 22 U.S. government agencies reported obligating foreign assistance in FY2015.

[9] For more on current GPRA requirements, see CRS Report R42379, *Changes to the Government Performance and Results Act (GPRA): Overview of the New Framework of Products and Processes*, by Clinton T. Brass.

[10] Foreign Assistance Act of 1961, P.L. 87-195), §101(a).

[11] Ibid.

[12] FAA, as amended, §481(a)(1)(C).

[13] FAA, as amended, §491(a).

ties of the United States with friendly governments by offering concrete [antiterrorism] assistance"[14] were added in 1983. In short, U.S. foreign aid is intended to be a tool for fighting poverty, enhancing bilateral relationships, and/or protecting U.S. security and commercial interests.

In this broad view, some instances of specific development assistance projects and programs are widely viewed as successful. The largest aid program of the last century, the Marshall Plan (1948-1952), for example, is acclaimed as a key factor in the post-World War II reconstruction of European states that have gone on to become major strategic and trade partners of the United States. In the late 1960s and 1970s, aid associated with the "green revolution" was credited with greatly improving agricultural productivity and addressing hunger and malnutrition in parts of Asia, and global health programs were credited with virtually eradicating smallpox. Korea, Taiwan, and Botswana are often cited as aid success stories as a result of remarkable economic progress following significant aid infusions. More recently, unquestionable progress in battling public health crises, such as HIV/AIDS, across the globe can be largely attributed to massive foreign assistance programs, both bilateral and multilateral. Recent studies have also shown a positive but modest impact of aid on economic growth rates.[15] Even in these instances, however, close analysis often reveals many caveats.

In other specific instances foreign aid programs and projects have been considered to be conspicuously unsuccessful, or even harmful to intended beneficiaries. Critics of foreign assistance cite decades of aid to corrupt governments in Africa, which enriched corrupt leaders and did little to improve the lives of the poor.[16] In Latin America, U.S. aid to anti-communist rebels and regimes during the Cold War was associated with brutal violence and believed by many to have damaged U.S. credibility as a champion of democracy. Numerous examples exist of hospitals, schools, and other facilities that were built with donor funds and left to rot, unused in developing countries that did not have the resources or will to maintain them. In some instances, critics assert that foreign aid may do more harm than good, by reducing recipient government accountability, fueling corruption, damaging export competitiveness, creating dependence, and undermining incentives for adequate taxation.[17]

The most notable successes and conspicuous failures of foreign aid give fodder to both aid advocates and detractors, but in all likelihood represent just a small segment of assistance activities. In most cases, clear evidence of the success or failure of U.S. assistance programs is lacking, both at the program level and in aggregate. One reason for this is that aid provided for development objectives is often conflated with aid provided for political and security purposes. Another reason is that historically, most foreign assistance programs are never evaluated for the purpose of determining their impact, either at the time of implementation or retrospectively. Furthermore, evaluation practices are not consistent enough to allow for the use of project level data as the basis for broader, strategic evaluations. A 2009 review of monitoring and evaluation of U.S. foreign assistance described the evaluation effort at that time as "uneven across agencies, rarely assesses impact, lacks sufficient rigor, and does not produce the necessary analysis to

[14] FAA, as amended, §572 (1) and (2).

[15] "The $138.5 Billion Question: When Does Aid Work (And When Doesn't It)?," Center for Global Development Policy Paper 049, Sect. 3.1.

[16] Several examples of this are discussed in, *Economic Gangsters: Corruption, Violence and the Poverty of Nations*, by Raymond Fisman and Edward Miguel, Princeton University Press, 2008.

[17] See Dambisa Moyo, *Dead Aid: Why Aid is Not Working and How There Is a Better Way for Africa*, Farrar, Straus and Giroux, New York, 2009, p. 48.

inform strategic decision making."[18] In recent years, however, aid-implementing agencies have taken steps to improve both the quantity and quality of aid evaluations, and to make better use of the information gleaned from those efforts. A 2016 USAID review identified notable improvements in evaluation practices at USAID since implementation of a new evaluation policy in 2011.[19]

Impact and Performance Evaluations

The Department of State, USAID, and other U.S. agencies implementing foreign assistance programs consistently monitor the performance of their own personnel and contractors in meeting discrete objectives, tracking project inputs and outputs. Depending on the nature of the project or program, staff and contractors might monitor the miles of road built, number of police officers trained, or changes in the use of fertilizers by farmers. These results can be compared to the initial program goals and expectations to determine whether the project or contract has been performed successfully. This type of oversight is called *performance monitoring*. Financial audits by agency Inspectors General, which examine whether funds are being used as intended, are also a common form of performance monitoring, particularly at the State Department. These audits are in addition to regular financial audits required by agencies of contractors, aid-implementing partners, and host government entities.

If the data gathered through performance monitoring are analyzed in an effort to explain how and why a program meets or fails to meet strategic objectives, this is called *performance evaluation*. Performance evaluations have typically been carried out sporadically, to address questions of efficiency, effectiveness, and sustainability, among other things. Performance evaluations represent the vast majority of foreign aid evaluations.

Performance monitoring and evaluation play an important part in project management but do little to answer questions about foreign aid effectiveness. Addressing this question, some argue, requires *impact evaluations*. Impact evaluations look not at the output of an activity, but rather at its impact on a development objective. For example, while performance monitoring of an education program may involve tracking the number of textbooks provided and teachers trained, an impact evaluation may determine how or if literacy or math skills had improved for the target group as compared to a similar group that did not receive the textbooks or teacher training. A performance evaluation of an HIV prevention project may report the number of public awareness events held or condoms distributed, and analyze this data in the context of program goals, while an impact evaluation of the same program would monitor changes in the HIV/AIDS infection rate of the targeted population relative to a control group. An impact evaluation of a police training program would look at the program's impact on civil order and public safety rather than simply report how many officers were trained or the value of equipment supplied. Impact evaluation can take many forms, ideally using a defined counterfactual, or control group, and baseline data to measure change that can be attributed to an aid intervention.[20] Randomized controlled trials, in which beneficiaries are randomly selected from a prequalified group and compared before and

[18] *Beyond Success Stories: Monitoring and Evaluation For Foreign Assistance Results, Evaluator Views of Current Practice and Recommendations for Change*, by Richard Blue, Cynthia Clapp-Wincek and Holly Benner, May 2009, p. ii.

[19] "Strengthening Evidence Based Development: Five Years of Better Evaluation Practices at USAID, 2011-2016," available at https://www.usaid.gov/sites/default/files/documents/1870/Strengthening%20Evidence-Based%20Development%20-%20Five%20Years%20of%20Better%20Evaluation%20Practice%20at%20USAID.pdf.

[20] For a thorough, yet nontechnical, discussion of the use of impact/attribution evaluation, see "*An introduction to the use of randomized control trials to evaluate development interventions*," by Howard White, International Initiative for Impact Evaluation, Working Paper 9, February 2011.

after the program to those not selected, are widely viewed as best practice for impact evaluation, but less rigorous methods are used as well. For example, "before and after" data analysis, case studies, and mixed method designs using both qualitative and quantitative data may be used for impact evaluation.

Impact evaluations can be a key to determining whether a foreign assistance program "works." However, impact evaluations are generally far more complex and resource-intensive than performance monitoring and evaluation, and usually must be planned before an activity begins. Agencies implementing foreign assistance must balance the potential knowledge to be gained from impact evaluation with the additional resources necessary to carry out such evaluations. As a result, while the potential learning benefits of impact evaluation have long been recognized by aid officials, the use of rigorous impact evaluation has been, and continues to be, very limited. More typically, agencies aim for evaluation practices that are, as one expert has put it, "cost-effectively rigorous," and, at minimum, "independent, transparent, and consistent, thus persuasive."[21]

History of U.S. Foreign Assistance Evaluation

The practice of foreign assistance evaluation has changed over time to reflect evolving, or some might say cyclical, attitudes about the purpose and relative importance of evaluation.[24] This is evident both in the United States and internationally. Aid evaluation practices and policies have variously focused on different evaluation objectives, including meeting program management needs, institutional learning, accountability for resources, informing policymakers, and building local oversight and project design capacity.

The history of U.S. foreign assistance evaluation begins with USAID, which implemented the vast majority of U.S. foreign assistance prior to the last decade. In its early years, USAID was primarily involved in large capital and infrastructure projects, for which evaluations focused on financial and economic rates of return were appropriate. However, the agency soon shifted focus towards smaller and more diverse projects to address basic human needs, and found that the rate of return

Primary School Deworming in Kenya (1997-2001)[22]

One well-known example of an impact evaluation that yielded useful information looked at a World Bank-supported project in Kenya that treated children for intestinal worms, a prevalent affliction that results in listlessness, diarrhea, abdominal pain, and anemia. The stated development objective was to increase the number of children completing their primary education. In collaboration with the local health ministry, NGO implementers treated 30,000 children in 75 schools with a drug that cost $3.27 annually per child, using baseline data and a random phase-in approach that allowed for a controlled comparison. The evaluation found that the deworming resulted in a 25% reduction in absenteeism, or 10-15 more days of school attendance per child per year. This case is also an example of the value of consistent methodology and the use of sector- or region-wide evaluation that looks at results beyond the project level. Similar evaluation methods were used for other interventions (providing free uniforms, textbooks, and/or meals) with the same goal and in the same region, allowing evaluators to do a comparative analysis and determine that the deworming intervention was the most effective of these interventions in increasing school participation.[23]

[21] Clemens, Michael. "Impact Evaluation in Aid: What For? How Rigorous?" Presentation at the Overseas Development Institute, July 3, 2012, video recording available at http://www.cgdev.org/content/multimedia/detail/1426372/.

[22] For an overview of this evaluation, as well as links to related studies, see http://www.povertyactionlab.org/evaluation/primary-school-deworming-kenya.

[23] Roetman, Eric. *A Can of Worms? Implications of Rigorous Impact Evaluations for Development Agencies*, International Initiative for Impact Evaluations, Working Paper 11, March 2011, p. 5.

[24] Trends in Development Evaluation Theory, Policies and Practices, USAID, 17 August 2009, p. 4.

evaluation model was no longer sufficient.[25] The agency established its first Office of Evaluation in 1968, and used a Logical Framework (LogFrame) model as its primary system for monitoring and evaluation.[26] The LogFrame approach, subsequently adopted by many international development agencies, employed a matrix to identify project goals, purposes, results, and activities, with corresponding indicators, verification methods, and important assumptions. Baseline data were to be used for each indicator, and results were reported at quarterly points during the life of a project. However, these data were not analyzed to look for competing explanations of the results or unintended consequences of activities. In many respects, the LogFrame approach was quite similar to the current GPRA requirements (discussed in the "Program Evaluation Government-Wide" text box above.)

While the LogFrame approach established USAID as a thought leader with respect to evaluation policy, in practice, evaluation quality varied significantly from project to project. A 1970 evaluation handbook included a diagram of the "ideal" program evaluation design, which resembles a randomized controlled trial, but notes that "there are a great many reasons why it may not be possible to reach the ideal."[28] Reviews of foreign assistance evaluation over decades revealed shortcomings. For one, the system had become decentralized over time, suitable to meet the information needs of project managers in the field but not contribute to broader learning or policy making. A 1982 report by the General Accounting Office (now the Government Accountability Office, GAO) found that "AID staff does not apply lessons learned in the development of new projects," and that "lessons learned are neither systematically nor comprehensively identified or recorded by those who are directly involved."[29] In response to the GAO report's recommendation that USAID build an "information analysis capability," the agency created the Center for Development

Testing Family Planning Project Design in Thailand, 1979

Many evaluations are designed to answer specific questions about project design. One example is the Family Planning Health and Hygiene Project, a 1979 independent evaluation of USAID support for the government of Thailand's family planning policy. Implemented by the American Public Health Association, the evaluation used a baseline survey and experimental design to test the hypothesis that contraception services would be more cost-effective and acceptable to communities if combined with basic health services rather than implemented in isolation. Obtaining the appropriate information to inform resource allocation was a primary objective of the evaluation. According to the report, "the evaluation was implemented with sufficient precision and adherence to experimental requirements to provide information on which to make management decisions about the best use of resources." Evaluators found that the hypothesis was not supported by the evidence. Adding basic health services doubled the cost of programs but was not associated with increased contraceptive use. As a result, the evaluators recommended that future decisions about family planning and basic health services programs be considered without any assumption that a linkage between the two would increase the acceptance of contraception use.[27]

[25] *The USAID Evaluation System: Past Performance and Future Direction*, Bureau for Program and Policy Coordination, USAID, September 1990, p. 9.

[26] That same year, the Foreign Assistance Act of 1961 (P.L. 87-195) was amended by the Foreign Assistance Act of 1968 (P.L. 90-554) to add Section 621A, which calls for "strengthened management practices," including defined objectives, quantitative indicators of progress, and means for comparing anticipated results with actual results.

[27] *The Community-Based Family Planning Services Family Planning Health and Hygiene Project*, prepared by Bruce Carlson, MSPH, and Malcolm Potts, M.D. under the auspices of The American Public Health Association, USAID, 1979, pp. 5, 7.

[28] *Evaluation Handbook*, Office of Program Evaluation, USAID, November 1970, p. 40.

[29] *Experience – A Potential Tool for Improving U.S. Assistance Abroad*, U.S. Government Accountability Office, GAO-ID-82-36, June 15, 1982, p. i (summary).

Information and Evaluation (CDIE) in 1983, with a mandate to "foster the use of development information in support of AID's assistance efforts."[30] CDIE carried out meta-evaluations to reveal broader trends in aid impact, provided information and training on evaluation best practices to mission staff, and made a wide range of evaluation reports accessible to implementers in the field. Aid officials suggest that CDIE's evaluation work played a significant role in shaping USAID strategies and priorities in many sectors over decades.

An internal USAID review in 1988 found that CDIE had greatly increased the use of aid evaluation information by implementers, but also identified a need to improve the quality and timeliness of evaluation reports.[31] While the evaluation policy at the time still called for rigorous, statistical methods of evaluation, it was found that this approach was never actually widely used at USAID because the required skills, time, and expense made implementation difficult.[32] As one internal review noted, "statistical rigor in evaluation methods was deemphasized in favor of 'reasonably' valid evidence about project performance."[33] Guidance to missions encouraged the use of low-cost and timely qualitative evaluation methodologies, including the use of key informant interviews, focus group discussions, community meetings, and informal surveys.[34]

In the early 1990s, accountability for funds became a primary focus of aid evaluation. After a 1990 GAO review concluded that USAID evaluation practices made it difficult or impossible to account for use of aid funds,[35] attention turned to tracking where aid money was going, not measuring what it was accomplishing. At the same time, USAID was facing increasing budgetary pressure and increasing congressional and public concern about what was being achieved through foreign assistance.[36] In response, USAID carried out an Evaluation Initiative from 1990 to 1992, greatly expanding the staff and budget of CDIE and making significant investments in rigorous evaluation designs and innovative methods to evaluate sector-wide results.[37] However, by the mid-1990s the priorities changed once again. A 1993 agency reorganization led to the 1994 elimination of an Office of Evaluation within CDIE, a reduction of overall CDIE staff,[38] and a new emphasis on "rapid appraisal techniques," which guidance documents describe as a compromise between slow, costly, and credible formal evaluation methods and cheap, quick, informal methods (focus group, etc.) that may be less reliable.[39]

In 1995, USAID replaced the requirement to conduct mid-term and final evaluations of all projects with a policy calling for evaluation only when necessary to address a specific management question.[40] The rationale was that the required evaluations had become pro forma, as

---

[30] *The History of CDIE*, CDIEHIST.017/SESmith;JREriksson/10-17-94, p.4.; available through the Development Experience Clearinghouse on the USAID website.

[31] Ibid.

[32] *The A.I.D. Evaluation System: Past Performance and Future Directions*, Bureau for Program and Policy Coordination, Agency for International Development, September 1990, p. 10.

[33] Ibid., p. 11.

[34] Ibid., p. 11.

[35] *Accountability and Control Over Foreign Assistance*, GAO/T-NSIAD-90-25, March 29, 1990, p. 6, 11. The review found that military assistance managed by State and the Department of Defense was also inadequately monitored and accounted for.

[36] *The History of CDIE*, p.6; *The A.I.D. Evaluation System*, p. 11.

[37] Ibid., pp. 6-7.

[38] Ibid., p. 8.

[39] *The Role of Evaluation in USAID*, Performance Monitoring and Evaluation TIPS, USAID CDIE, 1997, Number 11, p. 3.

[40] *Beyond Success Stories*, p.7; *Evaluation of Recent USAID Evaluation Experience*, Cynthia Clapp-Wincek and

GAO reviews had suggested, and that fewer, more comprehensive evaluations would be a better use of time and resources. As a result, the number of completed evaluations dropped from 425 in 1993 to an estimated 138 in 1999,[41] but the depth and scope of new evaluations reportedly did not change.[42] One study suggests that inconsistent guidance on evaluation in these years allowed many already overburdened mission staff to ignore agency-wide requirements, but noted that the Global Health, Africa, and Europe & Eurasia bureaus, which had their own evaluation procedures, continued to carry out quality evaluation work.[43]

Foreign assistance levels grew rapidly starting in 2003 to support military activities in Afghanistan and Iraq, as well as the President's Emergency Plan for AIDS Relief (PEPFAR) and the creation in 2004 of the Millennium Challenge Corporation (MCC). Accountability to Congress became a major evaluation priority. In 2005, inspired by remarks made by then House Foreign Operations Appropriations Subcommittee Chairman Jim Kolbe regarding the importance of being able to clearly demonstrate results of aid expenditures, USAID Administrator Andrew Natsios sought to revitalize evaluation within the agency. He sent a cable to all mission directors calling for the inclusion of evaluation plans, and higher quality evaluations, in all program designs; designated monitoring and evaluation officers at each post; and set aside funding for evaluations and incentives for employees who do evaluations; among other things.[44]

In 2006, in further pursuit of accountability, as well as a desire to rationalize the bilateral assistance efforts of multiple U.S. agencies, Secretary of State Condoleezza Rice created the Office of the Director of Foreign Assistance (F Bureau) at the State Department. In addition to consolidating many USAID and State policy and planning functions for foreign assistance, the F Bureau established an extensive set of standard performance indicators "to measure both what is being accomplished with U.S. Government foreign assistance funds and the collective impact of foreign and host-government efforts to advance country development."[45] Prior to this initiative, the State Department, which traditionally had managed a much smaller aid portfolio than USAID, is said to have made a de facto decision not to evaluate its assistance programs on a systematic basis.[46] The data collected through the "F process," which remains in place today, allow for a marked improvement in aid transparency, demonstrating comprehensively where and for what purpose aid funds are allocated by State and USAID as of FY2006.[47] However, the demands of F

Richard Blue, Working Paper No. 320, U.S. Agency for International Development, Center for Development Information and Evaluation, June 2001, p. 31.

[41] *Evaluation of Recent USAID Evaluation Experience*, p. 5. The report authors note that while some of the declining numbers can be attributed to missions not submitting their evaluations to the Development Experience Clearinghouse, as policy required, making the specific numbers unreliable, the trend of decline is unmistakable.

[42] *Evaluation of Recent USAID Evaluation Experiences*, p. 12.

[43] *The Evaluation of USAID's Evaluation Function: Recommendations for Reinvigorating the Evaluation Culture Within the Agency*, Janice M. Weber, Bureau for Program and Policy Coordination, USAID, September 2004, pp. 5, 10.

[44] *Actions Required to Implement the Initiative to Revitalize Evaluation in the Agency*, UNCLAS STATE 127594, July 8, 2005.

[45] See http://www.state.gov/f/indicators/index.htm. It was originally expected by many that the F Bureau would eventually track all foreign assistance provided by U.S. agencies, not just State and USAID. As of 2012, some MCC data has been added to the Bureau's public database (www.foreignassistance.gov), but there does not appear to be momentum toward any expansion of F Bureau authority.

[46] *Beyond Success Stories*, p. 14. The State Department traditionally has used a variety of resources for monitoring its foreign assistance programs, including Mission and Bureau Strategic Plans, annual performance and accountability reports, and Office of Inspector General and Government Accountability Office reports, but had no systematic evaluation process (*Department of State Program Evaluation Plan*, FY2007-2012 Department of State and USAID Strategic Plan, Bureau of Resource Management, May 2007, Appendix II).

[47] The data is publically available at http://www.foreignassistance.gov.

process reporting were believed by some to have interfered with more results-oriented evaluation work at USAID, and a 2008 assessment of State's evaluation capacity found that several bureaus, including those that manage State's security assistance programs, still had little or no evaluation capacity.[48]

The structural reforms of the F Bureau came at a time of heightened congressional scrutiny of foreign aid. In 2004, Congress established the Helping to Enhance the Livelihood of People (HELP) Around the Globe Commission, through a provision in P.L. 108-199, to independently review foreign assistance policy decisions, delivery challenges, methodology, and measurement of results. After nearly two years of work, the HELP Commission released its report in late 2007. On the subject of evaluation, the report noted that "everyone to whom members of the Commission spoke about monitoring and evaluation expressed concern about the inadequacy of the existing process" and concluded that "unless our government better evaluates projects based on the outcomes they achieve, it will not improve the effectiveness of taxpayer dollars."[49] The commission recommended creation of a unified foreign assistance policy, budgeting, and evaluation system within State, quite similar to the F process, which was established before the report was released. Other HELP Commission recommendations included ensuring that evaluation strategies use control groups and randomization as much as possible; considering new evaluation methods, such as the use of professional associations or accreditation agencies; and building, in collaboration with other donors, the capacities of recipient governments to provide reliable baseline data.[50]

At the same time the F Bureau was established, and the HELP Commission was active, the international donor community began to prioritize aid effectiveness, sparking renewed interest in rigorous impact evaluation (see the "A Global Perspective on Aid Evaluation" text box below). Some aid professionals viewed the F process as an opportunity to build a cross-agency aid evaluation practice focused on impact, and were disappointed that the common indicators used by the F Bureau, while an improvement with respect to comparability, measured outputs rather than impact. Furthermore, the use of more rigorous evaluation methodologies was not a focus of the reform.

These issues were revisited by the Obama Administration when it embarked in 2009 on a Quadrennial Diplomacy and Development Review (QDDR) to examine how State and USAID could be better prepared for current and future challenges. As a result of that review, the Administration committed itself in December 2010 to several principles of foreign assistance effectiveness, including "focusing on outcomes and impact rather than inputs and outputs, and ensuring that the best available evidence informs program design and execution."[51] The first QDDR became the basis of many changes at State and USAID, including the creation of a new Office of Learning, Evaluation and Research at USAID and a new USAID evaluation policy, which took effect in January 2011.[52] A second QDDR, in 2015, called for training to deepen

[48] *Beyond Success Stories*, p. 8.

[49] *Beyond Foreign Assistance: The HELP Commission Report on Foreign Assistance Reform*, The United States Commission on Helping to Enhance the Livelihood of People (HELP) Around the Globe Commission, December 7, 2007, p. 15.

[50] HELP Report, p. 99.

[51] QDDR, p. 110.

[52] A second QDDR, completed in 2015, continues to emphasize the need for better evaluation practices, calling for a "data-driven, evidence-based" approach to development and diplomacy policymaking, increasing evaluation training and capacity building, and noting that State's Bureau for Political and Military Affairs is developing a comprehensive approach to monitoring and evaluating security assistance programs.

evaluation expertise at both USAID and State, and for adding "rigor" to evaluations through better use of diagnostics and data analysis.[53]

The State Department adopted an evaluation policy similar to that of USAID in February 2012, requiring all large projects and programs to be evaluated at least once in their lifetime or five-year period, all State bureaus to complete two to four evaluations before the end of 2012-2013, and posts to do the same in 2013-2014. The 2012 policy also called for 3%-5% of program resources to be identified for evaluation purposes. It appears, however, that some of these requirements were not met, and in January 2015, State revised its policy, paring it down to a less directive form that was thought to be more appropriate for the wide range of State activities, from diplomatic engagement to foreign assistance, and to reflect ongoing challenges in evaluating particularly sensitive activities such as security assistance (see the "Evaluating Security Assistance" text box below).[54] The new policy removed the requirement that all large projects be evaluated, requires one evaluation per bureau per year, and does not require any evaluations at the post level. Further details of the new policy are provided in the **Appendix**.

MCC Rural Water Supply Project in Mozambique, 2008-2013

One MCC impact evaluation looked at a rural water supply project that was part of the $507 million Mozambique compact that ended in 2013. The $200 million project installed water points (mostly hand pumps) in 614 poor, rural communities, with the expectation that better access to improved water sources would reduce waterborne disease rates and allow women and girls to spend less time fetching water and more time on education or economically productive activities. The program met or exceeded most of its output targets, which related to water points constructed, number of people trained in sanitary best practices, percentage of population with improved water access, and time saved to get to primary water source. From a performance perspective, it was a success. The independent impact evaluation, however, showed that improved access to clean water did not have any statistically significant impact on beneficiary health or income, which were the ultimate objectives. Analysis of the results revealed that while water quality was high at the collection point, it often became contaminated at the household level, possibly negating the health benefits of the improved water points. The evaluation did not discuss potential reasons why the average of an hour saved every day in water collection did not translate into higher household income. Nevertheless, this evaluation challenged assumptions on which the project was designed, offering significant learning value. In response to the evaluation findings, MCC reported that it would take steps to enhance peer review of critical assumptions, improve understanding of local community water sanitation knowledge and practices before designing future water supply projects, and consider how evaluators can assign value to time savings beyond income generation. Evaluators also suggested that a longer time frame may be necessary to observe income-related results, and MCC reports that it may conduct a survey in 2016 to assess the longer-term impacts of this project.

Source: *Measuring Results of the Mozambique Rural Water Supply Project*, MCC, August 11. 2014, available at https://www.mcc.gov/resources/doc/summary-measuring-results-of-the-mozambique-rwsa.

The Millennium Challenge Corporation, established in 2004, has been regarded by many as a leader in aid evaluation, largely as a result of its demanding evaluation policy. MCC provides funding and technical assistance to support five-year development plans, called "compacts," created and submitted by partner countries. Since its inception, MCC policy has required that every project in a compact be evaluated by independent evaluators, using pre-intervention baseline data. MCC has also put a stronger emphasis on impact evaluation than State and USAID; of the 48 completed evaluations as of April 2016, 13 are described as impact evaluations (as are about 40 of the 101 planned evaluations), a much high proportion than at other aid agencies.[55] Despite this emphasis, the overall impact of MCC assistance remains unclear. Individual project

[53] 2015 QDDR, pp. 13, 57.

[54] Conversations between CRS and State Department officials, February 2015, May 2016.

[55] This data was provided to CRS by MCC on April 15, 2016. Includes evaluations of both compacts and threshold programs.

evaluations have demonstrated successful project implementation, but often little evidence of progress toward the overarching objective of raising household incomes in targeted areas. Such evidence, however, may only be apparent many years after compact completion.

Evaluation Challenges

The current evaluation emphasis on measuring impact and broader learning about what works is not new; as discussed above, it was the basis of USAID evaluation policy in the 1970s and at various times since. Nevertheless, a 2009 meta-evaluation of U.S foreign aid programs indicated that rigorous impact evaluation—the kind that could determine with credibility whether a specific aid intervention or broader sector strategy worked to produce a specific development outcome—was rarely attempted. Of the 296 evaluations posted between 2005 and 2008 to USAID's Development Experience Clearinghouse website, an independent reviewer found only 9% reported on a comparison group and only one used an experimental design involving randomized assignment, the method most likely to produce accurate data.[56] A 2005 review of USAID evaluations (focused on democracy and governance programs) found that "as a group, they lacked information that is critical to demonstrating the results of USAID projects, let alone whether the projects were the real cause of whatever change the evaluation reported."[57] A meta-evaluation covering the period 2009-2012 found a notable increase in evaluation following the new evaluation policy and found improvements in 68% of quality factors examined, including the inclusion of recommendations. For most factors, however, the improvements were less than 15%, and most evaluations met USAID quality standards in only a few of the 37 criteria reviewed.[58] USAID anticipates completing a second meta-evaluation, covering the period 2012-2016, in 2017.

Evaluating Security Assistance

Foreign assistance evaluation efforts have focused almost exclusively on development assistance and, to a far lesser degree, humanitarian assistance. Military and security assistance programs under State Department authority have gone largely unevaluated. The strategic and diplomatic sensitivities of this type of aid present significant challenges for evaluators. Past efforts by State to contract independent evaluators for these programs were reportedly unsuccessful, with the unprecedented nature of the work creating high levels of uncertainty and perceived risk among potential bidders. These challenges may be one reason that State loosened its evaluation requirements in 2015 and why proposed legislation calling for more stringent and comprehensive aid evaluation has typically excluded security assistance. The 2015 QDDR, however, noted that the State Department's Bureau of Political-Military Affairs was developing a comprehensive approach to monitoring and evaluation of security assistance. A working group is reportedly tasked with establishing a feasible, incremental approach to security assistance evaluation, starting with the limited collection of baseline data. Initiate pilot evaluations of Foreign Military Financing programs may occur as early as 2017.

Sources: 2015 QDDR, p. 34; CRS conversations with State Department officials.

The gap between evaluation goals and actual practices has been documented repeatedly over the history of U.S. foreign assistance. So, too, have the challenges that make it difficult for implementers to achieve ideal evaluation practices in the field. Some of these challenges are discussed below.

[56] Trends in Development Evaluation Theory, Policies and Practices, USAID, 17 August 2009, p. 46.

[57] *Trends in International Development Evaluation Theory, Policies and Practices*; USAID, 17 August 2009, p. 13. The report was prepared for USAID by Molly Hageboeck of Management Systems International.

[58] A summary of the 2009-2012 meta-evaluation is available at http://usaidlearninglab.org/sites/default/files/resource/files/Meta%20Evaluation%20Presentation.pdf.

Mixed Objectives. The U.S. foreign assistance program has dozens of official objectives written into statute, and many aid programs are designed to meet multiple objectives. Often there are both strategic objectives and development objectives attached to an aid intervention, which may or may not be acknowledged in budget and planning documents. For example, assistance to Uzbekistan may have been requested and appropriated for specific agriculture sector activities, but may have been motivated primarily by a desire to secure U.S. overflight privileges for military aircraft bringing troops and supplies to Afghanistan. An evaluation of the agricultural impact may be of no use to policymakers who are more interested in the strategic goal, nor to aid professionals who are unlikely to view any lessons learned in these circumstances as applicable to agricultural development projects if political needs overrode the development rationale for the program.

Another example is the Food for Peace program, which provides U.S. agricultural commodities to countries facing food insecurity. One objective of the program is to feed hungry people, but long-standing requirements that most of the food be provided by U.S. agribusiness and be shipped by U.S.-flagged vessels make clear that supporting the U.S. agriculture and shipping industries is a program objective as well, and a potentially conflicting one. Studies have shown that the buy and ship America provisions, as they are known, may lessen the hunger-alleviation impact of food aid by up to 40%.[59]

Despite the political and diplomatic considerations that arguably underlie the majority of foreign aid, evaluations that examine those strategic objectives are rare (or at least not publicly available). This may be understandable, as such evaluations would often be politically and diplomatically sensitive. Nevertheless, evaluation that focuses only on the development or humanitarian impact of a particular program or project, when broader strategic objectives are drivers of the aid, may largely miss the point. For example, a 2015 Mercy Corp evaluation of youth employment programs in Afghanistan (funded by the United Kingdom, not the United States) tested the assumption that a program to create economic opportunities for youth would promote stability by lessening participants' support for political violence. Contrary to expectation, the evaluation found that the employment, economic confidence, and business connections fostered by the program made participants more likely to express support for political violence.[60]

Funding and Personnel Constraints. The more rigorous and extensive an evaluation, the costlier it tends to be, both in funds and staff time. Impact evaluations are particularly costly and require specially trained implementers. Absent a directive from agency leadership, aid implementers are unlikely to make resources available for evaluation at the expense of other program components. As one internal USAID review explained, "since USAID's development professionals have limited staff, limited budget, and copious priorities, unfortunately, due to lack of training on the crucial role of evaluation in the development process, most have chosen to eliminate evaluation from their programs."[61] Competitive contracting plays a role as well. At a time when most program implementation is contracted out, and cost is a key factor in winning contract bids, some argue that there is little incentive to invest in the up-front costs, such as baseline surveys, of a well-designed evaluation plan in the absence of an enforced requirement.[62] As a result, ad hoc evaluations of limited scope and learning value—as one report describes it, the

[59] *The Developmental Effectiveness of Untied Aid*, OECD, p.1, available at http://www.oecd.org/dataoecd/5/22/41537529.pdf.

[60] "Does Youth Employment Build Stability?," Evidence From Impact Evaluation of Vocational Training in Afghanistan, Mercy Corps 2015.

[61] *An Evaluation of USAID's Evaluation Function*, p. 5.

[62] *Beyond Success Stories*, p. 16.

"do the best you can in three weeks" approach—often prevail by default.[63] "It is rare," according to one report, "that the resources provided for an evaluation are sufficient to develop and apply more rigorous research methods that would produce valid empirical evidence regarding outcomes and attributable impact."[64] While MCC has the benefit of compacts being fully funded up front, which may account in part for its more comprehensive evaluation practices, State and USAID cannot count on receiving requested project funding from year to year, creating a challenge for all aspect of program implementation, including evaluation.

Sometimes the limited resource is personnel, rather than funding. Past reviews of assistance evaluation repeatedly cite lack of trained evaluation personnel as a problem. USAID has tried to address this problem by training 1,600 staff in evaluation design and implementation since 2011 and producing a number of evaluation tools, publications, and webinars available to staff. USAID has also recently recruited monitoring and evaluation fellows, who are placed for six months to two years in offices that need additional expertise.[65] Another part of this effort is building strong relationships with other entities focused on aid evaluation, including aid agencies of other donor countries and the International Initiative for Impact Evaluation (3ie).[66] Some experts have suggested that greater emphasis on collective evaluations—donor countries and foundations contributing to an independent organization that conducts evaluations of aid crossing many donor portfolios—could address resource and expertise limitations as well as allow for generalization of evaluation findings and policy relevance.[67]

Emphasis on Accountability of Funds. Aid monitoring and evaluation efforts over the past decade have primarily focused on accountability of funds because that is what stakeholders, including Congress, generally ask about. Concerned about corruption and waste, bound by allocation limits, and required by law to report on various aspects of aid administration, implementing agencies have developed monitoring, evaluation, and data collection practices that are geared toward tracking where funds go and what they have purchased rather than the impact of funds on development or strategic objectives. For example, the F Bureau's Foreign Assistance Framework, launched in 2006, was created largely to address the information demands of stakeholders, who wanted more data on how aid funds are being spent. It worked, to the extent that it is now easier to find information on how much aid is being spent in a given year on counterterrorism activities in Kenya, for example, or on agricultural growth programs in Guatemala.[68] But little if any of the resulting data addresses the impact of aid programs.

Methodological Challenges. In the complex environment in which many aid projects are carried out, it can be challenging to employ high quality evaluation methods. U.S. agency policies allow for a variety of evaluation methods (see **Appendix**), acknowledging that the most rigorous methods are not always practical. Sometimes it is impossible to identify a comparable control group for an impact evaluation, or unethical to exclude people from a humanitarian intervention for the purpose of comparison. Sometimes the goals are intangible and cannot be accurately documented through metrics. For example, it may be much harder to measure the impact of

[63] Ibid.

[64] Ibid.

[65] *Strengthening Evidence Based Development*, p. 12

[66] For more information on 3ie, see the "A Global Perspective on Aid Evaluation" text box below.

[67] *The Future of Aid: Building Knowledge Collectively*, Center for Global Development Policy Paper 050, January 2015.

[68] Foreign aid data from FY2006-FY2012 estimates, sorted by recipient country, year, agency (only State, USAID and MCC), appropriations account, and objective is readily available through the "Foreign Assistance Dashboard" at http://www.foreignaid.gov.

programs such as the Middle East Partnership Initiative, designed to strengthen relationships, than to measure more concrete objectives, such as reducing malaria prevalence. This may be one reason why reviews have found that global health assistance has a stronger evaluation history than other aid sectors;[69] disease prevalence and mortality rates lend themselves to quantification better than military personnel attitudes towards human rights or the strength of civil society. Rigorous methodology can also limit program flexibility, as making program changes mid-course, in response to changed circumstances or early results, can compromise the evaluation design. Some MCC evaluation reports note that information gleaned from early project implementation resulted in mid-course changes that improved program logic but undermined impact evaluation plans.

Even when metrics and baselines are well established, it can still be very difficult to attribute impact to a specific U.S. aid intervention when such programs are often carried out in the context of a broader trade, investment, political, and multi-donor environment.[70] A 2016 SIGAR report, for example, notes that while USAID frequently cites improvements in Afghanistan's education sector among the highlights of U.S. reconstruction efforts, the agency is unable to establish a link between U.S. assistance and trends in the sector, in which many donors are active.[71] Also, some aid professionals see broader drawbacks to rigorous impact evaluation methods. Some assert that the use of randomized control groups, which generally require the use of independent evaluators, limits the participation of affected individuals and communities in project design. They argue that community participation in project planning and evaluation, which can lead to greater buy-in and local capacity building, is more valuable in the development context than high-quality evaluation findings.[72] Others counter that more participatory methodologies are often weakened by bias, and that it is unwise and even unethical to replicate programs, which may profoundly affect participants, without having properly evaluated them.[73]

Compressed Timelines. While development assistance, in particular, is recognized as a long-term endeavor, aid strategies can be trumped by political pressures, which can influence evaluation. In 2001, a USAID survey report stated that "the pattern found was that evaluation work responds to the more immediate pressures of the day."[74] Policymakers facing relatively short budget and election cycles do not always allow adequate time for programs to demonstrate their potential impact. Such pressures have only increased over the past 15 years, particularly in the politically charged environments of Iraq, Afghanistan, and Pakistan. As a Senate Foreign Relations Committee majority-staff report on aid to Afghanistan found, "the U.S. Government has strived for quick results to demonstrate to Afghans and Americans alike that we are making progress. Indeed, the constant demand for immediate results prevented the implementation of programs that could have met long-term goals and would now be bearing fruit."[75]

The type of evaluation necessary to determine whether aid has real impact is both hard to do and of limited use in a short-term context. Timelines are particularly restrictive for MCC, which

[69] *Beyond Success Stories*, p. 9.

[70] The QDDR states that "we know that in many cases the outcome-level results are not solely attributable to U.S. government investments and activities; we will focus on outcome-level progress in locations and subsectors where the U.S. government is concentrating support." (QDDR 2010, p. 104).

[71] SIGAR Education report 16-32-AR p. 16. The report also notes that the education data used by USAID is provided by the Afghan government and has not been independently verified.

[72] *A Can of Worms*, p. 8.; *Beyond Success Stories*, p. 17.

[73] *Improving Lives Through Impact Evaluation*, p. 15

[74] *Evaluation of Recent USAID Evaluation Experiences*, p. 26.

[75] S.Prt. 112-21, Evaluating U.S. Foreign Assistance to Afghanistan, June 8, 2011, p. 14.

originally intended to complete evaluations during the compact implementation period. This goal, which reflects broad support for limited timeframes on foreign assistance, was found not to be feasible during implementation of MCC's first compacts in Cape Verde and Honduras.[76] Baseline data and evaluation models can be rendered worthless if program timelines change. For example, an MCC evaluation of a farmer training program in Armenia found that the planned impact evaluation model—a phased roll-out—was compromised by a delay in implementing one component of the program and the five-year compact timeline.[77] The long-term impacts of aid may be the most significant in judging effectiveness, but are least likely to be evaluated.

Sector Evaluation Example: Trade Capacity Building

Many analysts have suggested that cross-country evaluations of aid for a specific sector may be more useful for shaping policy than the more common individual project evaluations. One example of this approach is an evaluation commissioned by USAID to look at the impact of 256 U.S. trade capacity building (TCB) assistance projects in 78 countries from 2002 to 2006. The United States obligated about $5 billion during this period for TCB activities, through several federal agencies, including assistance to help developing countries strengthen their public institutions and policies related to trade, as well as programs to make private industries more knowledgeable about and competitive in global markets. The evaluation was designed after the fact, making a randomized controlled trial unfeasible, and had to account for variations in reporting across projects. Much of the report highlights anecdotal examples of issues that could not be analyzed systematically as a result of inconsistent data collection methodologies across projects. However, using regression analysis, evaluators found a relationship suggesting that each additional $1 invested in U.S. aid (from all agencies) for TCB is associated with a $53 increase in the value of recipient country exports two years later. For TCB aid specifically managed by USAID, the relationship was $1 invested for $42 in increased exports. No similar association was found between TBC assistance and recipient country imports or foreign direct investment. While this evaluation's methodology was not sufficient to demonstrate actual aid impact or causation, its findings may be useful to policymakers in both demonstrating a correlation between TCB aid and export growth, as well as forming the basis of a discussion about the comparative advantages of various U.S. agencies in managing TCB aid.

Source: *From Aid to Trade: Delivering Results. A Cross-Country Evaluation of USAID Trade Capacity Building*, "Executive Summary," prepared for USAID by Molly Hageboeck of Management Systems International, November 24, 2010.

Country Ownership and Donor Coordination. The United States and other aid donor countries have made pledges to both coordinate their efforts and increase recipient country control, or "ownership," over the planning of aid projects and the management of aid funds. Country ownership is believed by many to increase the odds that positive results will be sustained over time both by ensuring aid projects are consistent with recipient priorities and by helping to build the budget and project management capacity of recipient country governments and nongovernmental organizations (NGOs) that administer the assistance. Donor coordination of assistance efforts is supposed to promote efficiency, ease administrative burdens on aid recipients, and avoid duplication, among other things. USAID, as part of its ongoing procurement reform process, aims to channel an increasing portion of contract and grant aid directly to governments and local organizations. However, greater country ownership, and the pooled funds that may result from donor coordination, generally means diminished donor control, and a lesser ability to evaluate how U.S. funds contributed to a particular outcome. Accountability concerns often greatly overshadow the learning aspects of evaluation in such a context, as Congress has expressed concern about the heightened potential for corruption and mismanagement when funds flow directly to recipient country institutions. A 2016 report of the Special Inspector General for Afghanistan Reconstruction (SIGAR), for example, notes that while an increasing portion of U.S.

[76] *Millennium Challenge Corporation: Compacts in Cape Verde and Honduras Achieved Reduced Target*, GAO-11-728, p. 33.

[77] *Measuring Results of the Armenia Farmer Training Investment,* October 23, 2012, p.4, available at http://www.mcc.gov/documents/reports/results-2012-002-1196-01-armenia-results-country-summary.pdf.

aid to Afghanistan is being provided through Afghan government ministries, these ministries struggle with staffing, technical skills, management, and accountability.[78]

Security. Over the past 15 years, a significant percentage of foreign aid has been allocated to countries where security concerns have presented major obstacles to implementing, monitoring and evaluating foreign aid. A 2012 evaluation of a USAID agricultural development program in rural Pakistan, for example, states "the operating environment for development projects has been especially testing in recent years in the presence of an insurgency and frequent targeted killings and kidnappings."[79] Development staff in Afghanistan and Iraq in particular have not always been able to safely visit project sites to verify that a structure has been built or supplies delivered, much less be out on the streets conducting the types of surveys that certain evaluations would normally call for. A 2011 USAID Inspector General report noted that more than half of performance audits in Iraq at that time indicated security concerns, and a 2016 SIGAR report noted that the drawdown of U.S. and coalition military personnel in Afghanistan, and the deteriorating security situation, made it difficult or impossible for civilian agency personnel to oversee projects first-hand.[80] Even in less hostile environments, security concerns can undermine evaluation quality. For example, a 2011 evaluation of Office of Transition Initiatives governance activities in Colombia noted that "security considerations limited to some degree the evaluation team's freedom to interview community members in project sites at will. This fact made it difficult to be certain that field research did not suffer from a form of sampling bias."[81] While security challenges may weigh against the use of aid in certain regions, the most insecure places are sometimes where the U.S. foreign policy interests are greatest, and policymakers must consider whether the risk of being unable to evaluate even the performance of an aid intervention is worth taking for other reasons.

Agency and Personal Incentives. Given discretion in the use and conduct of evaluations, observers have noted the inclination of foreign assistance officials to avoid formal evaluation for fear of drawing attention to the shortcomings of the programs on which they work. While agency staff are clearly interested in learning about program results, many are reportedly defensive about evaluation, concerned that evaluations identifying poor program results may have personal career implications, such as loss of control over a project, damage to professional reputation, budget cuts, or other potential career repercussions.[82] As explained by one USAID direct-hire in response to a 2001 survey, "if you don't ask [about results], you don't fail, and your budget isn't cut."[83] That same study revealed that staff felt more pressure to produce success stories than to produce balanced and rigorous evaluations, and that "professional staff do not see any Agency-wide incentive to advance learning through evaluations."[84] Few observers consider risk taking and accepting failure as a necessary component of learning to be hallmarks of USAID or State Department culture, but a shift in this attitude may be in progress. According to USAID Administrator Gayle Smith, there has been "a cultural shift from checking the box that everything

[78] *Challenges to Effective Oversight of Afghanistan Reconstruction grow as High-Risk Areas Persist*, SIGAR, 2/24/16, pp. 9-10.

[79] *United States Assistance to Balochistan Border Areas: Evaluation Report*, Prepared by Management Systems International for USAID, January 16, 2012, p. vi.

[80] SIGAR 2/16 report, p. 14.

[81] *USAID/OTI's Integrated Governance Response Program in Colombia*, Final Evaluation, prepared by Caroline Hartzell et al., April 2011, p. 7.

[82] *Evaluation of Recent USAID Evaluation Experiences*, p. 22.

[83] Ibid., p. 24.

[84] Ibid., pp. 26-27.

is fine to here's what we're learning and here's what happened."[85] Other experts have suggested that there remains a reluctance within USAID to hold staff responsible for poor evaluation practices.[86]

Evaluating Humanitarian Assistance

Humanitarian assistance can present unique evaluation challenges, and is evaluated less frequently than development assistance. Available evaluation reports show significant shortfalls in this area. For example, a 2015 evaluation report of a State Department Bureau of Population, Refugees and Migration (PRM)-funded program to boost employment skills and opportunities for refugees living in camps in Ethiopia, implemented by three partners under the auspices of the United Nations High Commissioner for Refugees (UNHCR), found anecdotal evidence of positive program impacts but little basis for assessing program effectiveness. Neither PRM nor UNHCR at the time required more than basic monitoring of program outputs (individuals trained), and implementers could provide no data on livelihood or education outcomes, which were the objective of the programs. This was due in part to no system being in place to collect the necessary data, and in part because the camp population was fluid and many program participants left the camp soon after participating in the program and were not tracked. Despite the many challenges, U.S. agencies and other donors are making efforts to improve evaluation of humanitarian aid. Among the priorities that emerged from the 2016 World Humanitarian Summit consultative process is development of a framework and mechanisms for better evaluating the quality and effectiveness of humanitarian assistance by all donors.

Source: *Evaluating the Effectiveness of Livelihood Programs for Refugees in Ethiopia*, U.S. Department of State, available at http://www.state.gov/documents/organization/252133.pdf.

Applying Evaluation Findings to Policy

A consistent theme in past reviews of foreign aid evaluation practices is that even when quality evaluation takes place, the resulting information and analysis are often not considered and applied beyond the immediate project management team. Evaluations are rarely designed or used to inform policy. Lack of faith in the quality of the evaluation, irregular dissemination practices, and resistance to criticism may all contribute to this problem, as does lack of time on the part of aid implementers and policymakers alike to read and digest evaluation reports. A 2009 survey of U.S. aid agencies found that "bureaucratic incentives do not support rigorous evaluation or use of findings," "evaluation reports are often too long or technical to be accessible to policymakers and agency leaders with limited time," and learning that takes place, if any, is "largely confined to the immediate operational unit that commissioned the evaluation."[87] The shift in recent decades towards the use of contractors and implementing partners for most project implementation, and most project evaluation, may also impact the learning process. As one report notes, "partner organizations are learning from the experience, but USAID is not," and most evaluation work does not circulate beyond the partner.[88]

Congress expressed some interest in this issue with the Initiating Foreign Assistance Reform Act of 2009 (H.R. 2139 in the 111th Congress, introduced by Representative Howard Berman), which called for "a process for applying the lessons learned and results from evaluation activities,

[85] USAID Administrator Gayle Smith at a forum on "Assessing the Impact of Foreign Assistance: The Role of Evaluation," the Brookings Institution, March 30, 2016. See http://www.brookings.edu/events/2016/03/30-impact-foreign-assistance.

[86] Ruth Levine, Global Development and Population Program Director, Hewlett Foundation, at a forum on "Assessing the Impact of Foreign Assistance: The Role of Evaluation," the Brookings Institution, March 30, 2016. See http://www.brookings.edu/events/2016/03/30-impact-foreign-assistance.

[87] *Beyond Success Stories*, p.iv.

[88] *Evaluation of Recent USAID Evaluation Experiences*, p. 27.

including the use and results of impact evaluation research, into future budgeting, planning, programming, design and implementation of such United States foreign assistance programs." The government-wide GPRA performance planning and assessment requirements mentioned earlier (see "Program Evaluation Government-Wide" text box above) also attempted to mandate better use of evaluation data in policymaking government-wide. Aid agencies have addressed this issue with renewed focus and mixed results. USAID reviewed the utilization of evaluation data over the first several years under its new policy and found that 90% of surveyed evaluation findings and recommendations had some impact on program-level decisionmaking, mostly for project design and modification.[89] USAID requires that its five-year Country Development Cooperation Strategies (CDCS) cite evidence as the basis of their development hypothesis, and 60% of the CDCS in 2015 cited evaluation reports as evidence. However, there is no USAID requirement that new policies draw on evaluation findings, and the study found little evidence linking evaluations to higher-level policy decisions.[90]

The learning aspect of evaluation relies heavily on agency culture, which may be shaped more by leadership than policy. The effective application of evaluation information depends also on the details of implementation, such as evaluation questions being based on the information needs of policymakers and program managers, and information being presented in a format and to a scale that is useful. Policymakers, for example, may be much better able to make actionable use of a meta-evaluation of microfinance programs, presented in a short report highlighting key findings, than a whole database of detailed analysis of single projects, the results of which may or may not be more broadly applicable. Experts have pointed out that individual project evaluations, even when well done, do not roll up nicely into a document showing what works and what does not. They contend that for maximum learning, an effort must be made at the cross-agency or even whole-of-government level to develop evaluation meta-data that is responsive not only to the needs of a project manager interested in the impact of a particular activity, but also to agency leadership and policymakers who want to know, more broadly, what foreign assistance is most effective.

This view has been reflected in legislation introduced in recent Congresses. The Foreign Assistance Revitalization and Accountability Act of 2009 (S. 1524 in the 111th Congress, introduced by then Senator Kerry) called for the creation of a Council on Research and Evaluation of Foreign Policy to do cross-agency evaluation of aid programs. The Foreign Aid Transparency and Accountability Act (introduced in successive congresses by Senator Marco Rubio and Representative Ted Poe before being enacted and signed into law in July 2016), directs the President to establish guidelines for the consistent evaluation of foreign assistance across federal agencies.

As important as evaluation can be to improving aid effectiveness, not every aid project has broad learning potential. Knowing which potential evaluations could have the greatest policy implications may be key to maximizing evaluation resources. Many USAID projects, for example, are designed with no intention that they be scaled up or replicated elsewhere. In other situations, an approach may have already been well proven. In such instances, a basic performance evaluation for accountability may be appropriate, but rigorous evaluation may be a poor use of resources. A 2012 USAID "Decision Tree for Selecting the Evaluation Design" asks staff to first consider whether an evaluation is needed, and decline to evaluate if the timing is not

[89] *Evaluation Utilization at USAID*, February 23, 2016, p. 10.

[90] Ibid., p. 12.

right, if there are no unanswered questions for the evaluation to address, or if there is no demand from stakeholders.[91]

Current Agency Evaluation Policies

The primary U.S. government agencies managing foreign assistance each have their own distinct evaluation policies, with varying degrees of specificity. The Quadrennial Diplomacy and Development Review (QDDR) report of December 2010 stated the intent that USAID would reclaim its leadership role with respect to international development evaluation and learning, and referenced a new USAID evaluation policy in the works to reflect the growing demand for results data and attempt to address some persistent evaluation challenges. That policy took effect January 2011. The State Department followed suit in February 2012 with a new evaluation policy that was similar in many respects to the USAID policy, and MCC updated its policy in May 2012. State then updated its policy again in early 2015, apparently paring down several requirements in the 2012 policy, though the 2015 QDDR reaffirmed the State Department's commitment to building evaluation capacity. The **Appendix** table compares key provisions of the current evaluation policies of USAID, State, and MCC.

The State and USAID policies share much in common, balancing the costs and expected gains from evaluation. For example, both require performance evaluations of all larger-than-average projects and experimental/pilot projects, but not all projects. The policies share an emphasis on accessibility of information, with provisions to promote consistent and timely dissemination of evaluation reports, though State only requires public dissemination of foreign assistance evaluations, and summaries rather than full reports. In their introductory language, both policies emphasize the learning benefits of evaluation, in addition to accountability. The USAID policy is notably more detailed than State's on many of the issues. The USAID policy establishes required features for evaluation reports, and specifies that evaluation questions be identified in the design phase of projects, issues which the State policy does not address. USAID states that most evaluations will be conducted by third party contractors or grantees, to promote independence, while State's policy does not require independent evaluators. While USAID suggests a target allocation of 3% of program funds for program evaluation, the State policy provides no such target and the guidance suggests that such a target may not be realistic. Perhaps most significantly, USAID's policy calls for impact evaluation whenever feasible, while the State policy sets a clear expectation that impact evaluation will be rare.

MCC's evaluation policy shares many elements of the State and USAID policies, but goes farther in many respects. MCC requires independent evaluations of all compact projects, using indicators and baselines established prior to project implementation. The agency has also made a practice of including a "lessons learned" section in its evaluation reports. It may be, however, that first-hand experience with the challenges of evaluation is bringing MCC policy and practice closer to that of USAID over time. MCC's 2012 policy revision adopts definitions from USAID's 2011 evaluation policy and includes a section on institutional learning. The update also appears to move closer to the USAID model with respect to impact evaluation, calling for impact evaluations "when their costs are warranted," whereas the previous iteration referred to independent impact evaluations as an "integral part" of MCC's focus on results.[92] The MCC policy still appears to have the strongest enforcement mechanism among the three agency policies, conditioning the release of quarterly

[91] *Decision Tree for Selecting the Evaluation Design*, USAID, June 2012, p. 1, available on USAID's Development Experience Clearinghouse website.

[92] *Policy for Monitoring and Evaluation of Compacts and Threshold Programs*, MCC, May 1, 2012, p.18; *Policy for Monitoring and Evaluation of Compacts and Threshold Programs*, MCC, May 12, 2009, p. 17.

disbursements on substantial compliance with the policy. USAID's policy, in contrast, calls only for occasional compliance audits, and State's policy does not address compliance at all.

While some experts have called for greater uniformity of evaluation practices across agencies to allow for comparative analysis, others view the differences in State, USAID, and MCC evaluation polices as reflecting the different experience, scope of work, and priorities of the agencies. USAID, with the largest and most diverse assistance portfolio among the agencies, and numerous small projects, may require a more flexible approach to evaluation than MCC, which is narrowly focused on economic growth and recipient government ownership. At State, foreign assistance is just one part of a broader portfolio (including diplomatic activities), potentially impacting what type and scope of evaluation is useful or possible. State is also responsible for many military and security assistance programs, which present unique challenges, as discussed in the "Evaluation Challenges" section above.

These current evaluation policies may represent a step towards improving knowledge of foreign assistance measures of effectiveness at the program or project level, and increasing transparency of the evaluation process. They do not, however, attempt to establish a systemic approach to aid evaluation that would make country-wide, sector-wide, or cross-agency evaluation or aid more feasible. They look similar to earlier initiatives to improve aid evaluation. Many aspects of the 2011 USAID policy, for example, are strikingly similar to the required actions called for in the 2005 cable to USAID missions (e.g., evaluation planning as part of all program designs, designated evaluation officers at each post, and set-aside evaluation funds). It may be too early to know whether this new multiagency initiative will have more real or lasting impact than its predecessors. A meta-evaluation examining USAID evaluations from 2009 to 2012 indicates that both the number and quality of evaluations increased significantly in that period, but most evaluations in 2012 still failed to meet evaluation standards.[93]

[93] Meta-Evaluation of Quality and Coverage of USAID Evaluations: 2009-2012, August 2013, p. 7.

A Global Perspective on Aid Evaluation

U.S. foreign assistance evaluation efforts have evolved in the context of a global movement by public and private aid donors to improve aid effectiveness, with improved evaluation practices as one of many strategies. Representatives of aid donor countries meet regularly under the auspices of the OECD Development Assistance Committee (DAC) to discuss evaluation practices, among other things, as a means of implementing the aid effectiveness agenda laid out in the 2005 Paris Declaration on Aid Effectiveness and the 2008 Accra Agenda for Action. A 2010 OECD/DAC survey and report on evaluation in the development agencies of major donor countries highlighted several issues that are common to U.S.-specific aid evaluation.[94] The report found a heavy reliance on measuring outputs, but also a trend toward measuring aid impact and larger strategic questions of development effectiveness. It identified new emphasis on dissemination of evaluation findings, and found that while bilateral aid agencies on average allocated 0.1% of their development assistance budget to evaluation, lack of human resources—people qualified to do rigorous impact evaluations, evaluations of direct budget support, or requiring specific language skills, in particular—presented a bigger obstacle to evaluation goals than did financial constraints.

Nongovernmental organizations have focused on evaluation in recent years, as well. In 2004, an Evaluation Gap Working Group was convened by the Center for Global Development with support from the Bill & Melinda Gates Foundation and the William and Flora Hewitt Foundation. The Working Group focused on why rigorous impact evaluations of development assistance were so rare. The resulting report, "When Will We Ever Learn?," is a key resource for this report. The group made two recommendations: (1) that donors invest more in their own evaluation capacity, and (2) that an independent institution be created to evaluate aid.[95] The offshoot of the latter recommendation is the International Initiative for Impact Evaluation (3ie), established in 2009, with a mission to use impact evaluations, specifically, to generate high quality evidence for use in shaping effective development policies. 3ie both funds evaluations and produces extensive materials on evaluation methods, implementation practices, and application to policy, as a means to improve evaluators' technical capacity. USAID and MCC are official partners of 3ie, as are many other official aid agencies, private foundations, and nonprofit organizations such as the Hewlett and Gates foundations and Save the Children.

Issues for Congress

While some momentum on foreign aid evaluation reform has originated within the Administration, Congress may have significant influence on this process. Not only can Congress mandate or promote a certain approach to evaluation directly through legislation, as has been proposed, it can modulate Administration policies by controlling the appropriations necessary to implement the policies. Congress may also influence how, or if, the information resulting from evaluations will impact foreign assistance policy priorities. These issues are discussed in greater detail below.

Reform Authorization Legislation. In the 112th and 113th Congresses, legislation was introduced that focused specifically on foreign aid evaluation. The Foreign Aid Transparency and Accountability Act (H.R. 3159/ S. 3310 in the 112th, S. 1271/H.R. 2638 in the 113th Congress) sought to evaluate the performance of U.S. foreign assistance programs and improve program effectiveness by requiring the President to establish guidelines on measurable goals, performance metrics, and monitoring and evaluation plans for foreign assistance programs that can be applied on a consistent basis across implementing agencies.[96] The legislation also called for the creation of a website that would make detailed, program-level information on foreign assistance, including country strategies, budget documents, budget justifications, actual expenditures, and program

[94] *Evaluation in Development Agencies*, Better Aid, OECD Publishing, 2010, available at http://dx.doi.org/10.1787/9789264094857-en.

[95] *When Will We Ever Learn?: Improving Lives Through Impact Evaluation*, Report of the Evaluation Working Group, Center for Global Development, May 2006.

[96] The House and Senate proposals were similar but not identical. For example, H.R. 3159, as passed by the House, called for evaluation guidelines to be applied "with reasonable consistency," while S. 3310 called for the guidelines to be applied "on a uniform basis."

reports and evaluations available to the public. The legislation was reintroduced in the 114th Congress (H.R. 3766/S. 2184) with some modifications, including the exclusion of most security assistance. It was enacted and signed into law in July 2016 as P.L. 114-191, potentially shaping aid evaluation practices in the years to come.

The general focus of these proposals is on codifying evaluation requirements and extending them across the various federal and agencies that administer aid programs. The benefit of such broad uniformity, arguably, is that it could enable policymakers, the public, and other stakeholders to better compare the activities of various agencies and get a more comprehensive picture of total U.S. foreign assistance. A potential drawback is the effort and expense required to impose such uniformity on agencies with different objectives, management structures, and information technology systems. These proposals also focus on transparency and accountability rather than effectiveness, and do not explicitly promote the use of impact evaluation, though they call for the use of rigorous methodologies, including impact evaluation. If performance evaluation continues to comprise the vast majority of aid evaluations, such a cross-agency requirement may provide comparable information on aid management from agency to agency, but is not likely to facilitate comparative analysis of what aid channels are most effective.

Appropriations for Enhanced Evaluation. Increasing the number and quality of foreign aid evaluations, while potentially cost effective in the long run, requires an investment of resources. For the most part, evaluation costs are integrated into program accounts at the various implementing agency budgets and are not scrutinized specifically by Congress. Annual funding levels established by Congress, together with any related legislative directives that limit the use of funds, may play a role in determining the extent of the Administration's efforts and capacity to strengthen evaluation practice. Congress may also wish to specify in appropriations legislation a portion of funds to be used for evaluation purposes.

Impact of Evidence-Based Approach on Congressional Priorities. Congress has long exerted control over foreign assistance not only through appropriated funds and restrictions, but also by directing foreign assistance funds to certain sectors, countries, or even specific projects through bill or report language. For example, the committee reports accompanying the annual State-Foreign Operations appropriation proposals provide specific funding levels for microfinance, basic education, water and sanitation, women's leadership training, people-to-people reconciliation programs in the Middle East, and other sectors of particular interest to Members of Congress. Should credible information about the relative effectiveness of these programs be made available as a result of improved evaluation practices, Congress can weigh the importance of the data, among other considerations, in establishing aid priorities. Some congressional directives on aid are less likely than others to be affected by evaluation results. The availability of actionable evaluation data may not result in a maximization of aid effectiveness, but may allow Congress to make more deliberate trade-offs between effectiveness and other objectives.

Conclusion

The primary U.S. agencies charged with implementing foreign assistance have made significant steps in the last several years to address ongoing deficiencies in evaluation practices that make it difficult to judge whether foreign assistance is achieving its various objectives. There is widespread agreement on the need for more consistent performance evaluation of aid programs. The value of rigorous impact evaluation is broadly recognized as well, though the agencies differ in their capabilities and aspirations in this respect. Past policies and evaluation reform efforts, however, have been similarly focused but not sustained in the face of persistent challenges, many of which remain today. Other reforms, such as the establishment of centralized evaluation

processes or the creation of an independent evaluation entity, have been proposed in legislation but not yet enacted. Growing emphasis in Congress and the Administration on results-based budgeting, as well as movement within the international aid donor community toward more rigorous aid evaluation practices, may provide the context for sustained progress. The 114th Congress continues to have opportunities to influence how U.S. foreign assistance is evaluated through legislative proposals, appropriations, and oversight activities.

Appendix. Select Aspects of Current USAID, State Department, and MCC Evaluation Policies

	USAID	State	MCC
Effective Date	January 2011	January 29, 2015	May 1, 2012
Responsible Personnel	PPL/LER responsible for system implementation, while missions and functional bureaus responsible for conducting evaluations. All Bureaus and operating units must designate an evaluation point of contact.	F oversees planning and implementation of foreign assistance evaluations, BP for diplomatic engagement evaluations. Each Bureau is responsible for conducting its own evaluations and must appoint a Bureau Evaluation Coordinator.	Primary lead is MCA (host country entity) M&E, with input from MCC M&E.
Evaluation Requirement	Operating units must conduct at least one performance evaluation of each project that equals or exceeds average project size. Projects involving an untested hypothesis or new approach, and that are anticipated to expand in scale or scope, will undergo an impact evaluation, if feasible. All evaluations will share certain basic features, including a full description of methodology; standardized recording and maintenance of records from evaluation; evaluation findings based on facts, evidence, and data, sex-disaggregated data; and an explanation of the limitations of the data. Key evaluation questions will be identified during the design phase of every project.	All programs/projects/activities greater than or equal to the median size (using dollar value or staff resources as the measure) for the Bureau must be evaluated at least once in their lifetime. All pilot programs must be evaluated before being replicated. Each Bureau or office should conduct at least one evaluation each fiscal year.	All Compacts and Threshold Agreements include monitoring and evaluation plans, which identify the evaluations to be conducted for each project, the key evaluation questions and methodologies, and the data collection strategies that will be used. Final evaluations are required for all projects in a Compact upon completion or termination; mid-term evaluations are discretionary. Selected indicators must have baselines established prior to the start of the corresponding activity.
Evaluation Type	Emphasis on quality evaluation methods and favoring random assignment/experimental methods for impact evaluations when feasible.	Evaluations should be based on verifiable data and information that have been gathered using the standards of professional evaluation organizations. According to the guidance, counterfactual data required for impact evaluation "cannot be collected for the overwhelming majority of the evaluations of management processes, delivery system and programs – unlike in other fields, control groups are not established when projects or programs are initiated at the Department. Even when data can be generated, the cost of collecting can be prohibitive."	Impact evaluations performed "when their costs are warranted by the expected accountability and learning."

	USAID	State	MCC
Evaluator Type	Policy states that most evaluations will be conducted by third party contractors or grantees managed by USAID, but evaluation teams may be composed primarily of USAID staff, led by an outside expert, when it is determined that this will facilitate institutional learning.	Suggests that evaluators should be "free from and pressure and/or bureaucratic interference," but does not require the use of outside evaluators. Bureaus and offices may conduct evaluations with their own staff as long as the staff have the appropriate training and experience and are not accountable to the managers of the program being evaluated.	Independent evaluators required for final evaluations of Compacts. Mid-term compact evaluations and final threshold program evaluations can be done independently or by MCC/MCA staff.
Funding Requirement	Recommends an average 3% of program budgets be dedicated specifically to *external* evaluation, distinct from monitoring. Resources for evaluation should be concentrated on large projects and those that are innovative or pilot approaches.	Calls for program managers to identify resources to conduct evaluations during program planning, but does not specify an amount or portion of funds to be used for evaluation, and the guidance suggests that the international standard of 3-5% of program costs is unrealistic.	Does not specify a portion of funds that should be used for evaluation.
Reporting Requirement	Public availability of evaluation reports and summaries, within 3 months of completion, on the Development Experience Clearinghouse website.	Bureaus and posts must post summaries of evaluation results internally, unless they are classified or sensitive but unclassified (SBU). Summaries of foreign assistance evaluations must be posted publicly on the F Bureau web page of the state.gov website.	MCAs must post their approved Compact M&E plans on their website. MCC and MCAs must "regularly" publish results information on their websites.
Compliance Enforcement	PPL/LER will organize occasional external technical audits of operating unit compliance with the policy.	No reference to compliance enforcement.	Substantial compliance required for approval of quarterly disbursements requested by recipient country.

Sources: *Policy for Monitoring and Evaluation of Compacts and Threshold Programs*, MCC, May 1, 2012; *Department of State Evaluation Policy*, Bureau of Resource Management, February 23, 2012; *Evaluation: Learning from Experience, USAID Evaluation Policy*, January 2011.

Notes: PPL/LER = USAID Office of Learning, Evaluation and Research; F Bureau = Office of Foreign Assistance Resources; RM = State Department Bureau of Resource Management; MCA = the Millennium Challenge Account implementing entity in each compact country; M&E = monitoring and evaluation. The information in the table refers only to what is in the actual evaluation policy document of each agency, as cited above. Information available outside of these documents, which may provide greater details about aspects of the policies, is not reflected here.

Author Information

Marian Leonardo Lawson
Analyst in Foreign Assistance

Disclaimer

This document was prepared by the Congressional Research Service (CRS). CRS serves as nonpartisan shared staff to congressional committees and Members of Congress. It operates solely at the behest of and under the direction of Congress. Information in a CRS Report should not be relied upon for purposes other than public understanding of information that has been provided by CRS to Members of Congress in connection with CRS's institutional role. CRS Reports, as a work of the United States Government, are not subject to copyright protection in the United States. Any CRS Report may be reproduced and distributed in its entirety without permission from CRS. However, as a CRS Report may include copyrighted images or material from a third party, you may need to obtain the permission of the copyright holder if you wish to copy or otherwise use copyrighted material.