2020

Beyond Good Intentions: New Legislation on Foreign Aid Effectiveness

Gergana Danailova-Trainor

James Filpi

Norman L. Greene

Salome Tsereteli-Stephens

Beyond Good Intentions: New Legislation on Foreign Aid Effectiveness[a]

GERGANA DANAILOVA-TRAINOR,[*] JAMES FILPI,[**]
NORMAN L. GREENE,[***] AND SALOME TSERETELI-STEPHENS[****]

---

[*] U.S. Government Accountability Office. danailovatrainorg@gao.gov. The opinions and views expressed in this article are the author's alone and are not intended to reflect GAO's institutional views. Gergana Danailova-Trainor is a Senior Economist with the U.S. Government Accountability Office. Since coming to GAO in 2005, Gergana provided technical expertise and has led teams in International Affairs and Trade (IAT) in assessing the results of programs and the effects of policies implemented by U.S. government agencies. She has worked on multiple engagements in most major IAT portfolio areas, including international security, foreign assistance, international trade, human rights, international sanctions, and foreign affairs management. More recently, she has applied her knowledge of program evaluation and MCC's results framework to a broader meta-evaluation of foreign assistance programs across key U.S. foreign assistance agencies. Gergana holds a Ph.D. in economics from the Johns Hopkins University.

[**] Department of Commerce. jamesfilpi@gmail.com. The opinions and views expressed in this article are the author's alone and are not intended to reflect the U.S. Department of Commerce's institutional views. James Filpi is a Senior Counsel with the Commercial Law Development Program (CLDP) in the Office of the General Counsel of the United States Department of Commerce. This article does not represent the positions of the U.S. Department of Commerce. Prior to public service, Filpi practiced law with the International Competition Group at the Washington, D.C. office of the law firm Goodwin Procter, LLP. He advised foreign and domestic clients regarding commercial transactions, including mergers and acquisitions, federal administrative procedures, corporate restructurings, banking, and consumer protection. Prior to his legal career, Filpi managed research and development programs for the Washington, D.C.-based strategic research firm, Advisory Board Company, and tracked commerce and trade legislation for a member of the United States Congress. He received his Juris Doctorate from Georgetown University Law Center where he was a staff member of the Georgetown International Environmental Law Review and spent a portion of his legal study focusing on transnational law at the University of Hong Kong. He also holds a Bachelor's of Arts in Psychology from the University of Virginia. Filpi is the author of "United States Foreign Assistance: Beyond Good Intentions and Toward Accountability," ILSA Journal of International & Comparative Law, Vol. 23:3 (2016) (available in English and Spanish).

[***] Schoeman Updike Kaufman & Gerber, LLP, New York, N.Y. ngreene@schoeman.com. Norman Greene has written, spoken, and programmed on rule of law and development, including judicial reform; judicial independence; the design of judicial selection systems and codes; the developing world, particularly North Africa (Morocco); foreign assistance; and counter-human trafficking. His articles have appeared in a number of journals, including the Denver University Law Review (international rule of law and development; U.S. judicial elections); West Virginia Law Review (judicial bias and the rule of law); ILSA Journal of International and Comparative Law (judicial reform in Morocco; counter-human trafficking (co-authored)); and the ABA's Africa Law Today (co-authored.) He belongs to a number of

I.   Executive Summary

   The Foreign Aid Transparency and Accountability Act (FATAA) of 2016[1] may be considered the most significant legislation in U.S. foreign aid in decades. The new law and the corresponding OMB and key foreign aid agencies' guidelines require providers to follow best practices in the monitoring and evaluation (M&E) of U.S government (USG) foreign aid.[2] A recent study conducted by the U.S. Government Accountability Office across the key USG foreign aid agencies identified a number of areas that needed improvement in the design, implementation, conclusions, and dissemination of foreign assistance evaluations.[3] FATAA and the relevant guidelines will require providers to address those areas and focus their reporting requirements on tangible outcomes and the impact of their programming. But these requirements pose a number of implementation challenges for smaller projects, smaller implementers, and certain interventions with long-term effects, such as the programs supporting the rule of law. In addition, because foreign aid funding implementers throughout the USG provide programming in conjunction with each other, programming provided by multiple, collaborating agencies must evaluate the outputs and outcomes of this joint assistance. While the guidelines require a delineation of roles and responsibilities of funding and implementing

professional associations on international law, including the N.Y.C. and American Bar Associations, the American Society of International Law, and the American Branch of the International Law Association, and practices law at the firm of Schoeman Updike Kaufman & Gerber, LLP, New York, N.Y. He graduated from Columbia College and New York University School of Law. The opinions and views expressed in this article are his own.
**** Creative Associates. salome.tsereteli@gmail.com. The opinions and views expressed in this article are the author's alone and are not intended to reflect the views of current or former employers. Salome Tsereteli-Stephens is the Senior Manager of Design, Monitoring, and Evaluation at Creative Associates. She has over fifteen years of experience in international aid and development, specializing in monitoring and evaluation. She has developed monitoring tools, trained staff, and conducted evaluations for various organizations in the U.S. and overseas. Her interests include mixed methods approaches for measuring democracy, governance, the rule of law, evaluation of policy and advocacy, and use of evaluation. Over the course of time, she has written and presented on these issues in the U.S. and abroad. Originally from Georgia, Tsereteli-Stephens holds an M.A. in law from Tbilisi State University and is fluent in several languages. She has been a member of the American Evaluation Association since 2012 and co-chaired its Democracy and Governance Topical Interest Group in 2015 through 2017.

   1. Foreign Aid Transparency and Accountability Act of 2016, Pub. L. No.114-191, 130 Stat. 666.
   2. Off. of Mgmt. & Budget, Exec. Off. of the President, OMB Bull. No. 18-04, Monitoring and Evaluation Guidelines for Federal Departments and Agencies that Administer United States Foreign Assistance (2018) [hereinafter OMB Bull. No. 18-04] (Introductory Statement), https://www.whitehouse.gov/wp-content/uploads/2017/11/M-18-04-Final.pdf.
   3. U.S. Gov't Accountability Off., GAO-17-316, Foreign Assistance: Agencies Can Improve the Quality and Dissemination of Program Evaluations 1 (2017) [hereinafter GAO-17-316], https://www.gao.gov/assets/690/683157.pdf.

agencies with regard to M&E, various aspects of interagency coordination require further clarification: program definition, logic model development collaboration, outcome attribution to individual agencies, coordinated monitoring and evaluation frameworks in joint program implementation, and reporting to OMB.

II.   Context and Focus of the FATAA Legislation and Agencies' Responses

The key strategic goal of U.S. foreign aid is to improve the lives of people living in poorer countries by shifting resources from richer to poorer countries. This is unlike the colonial era when resources were taken from poorer nations to benefit the richer ones. The scope of foreign assistance covers a broad spectrum of interventions: human rights, education, health, and overall economic development. Assessing foreign aid program results and improving aid delivery can help development professionals determine what works well. There are many reasons why some aid programs do not work well. In some cases, the conditions for development do not exist due to lack of political will for reforms in receiving countries or due to local conditions in which corruption may be siphoning off aid.[4] In other cases, programs are not designed appropriately or there arise unintended consequences. One way of improving aid effectiveness is through better evaluation requirements for development programs.

A.   IMPETUS AND CONTEXT OF FATAA

In theory, M&E are essential tools for U.S. agencies to assess and improve program results.[5] But evaluations require resources to perform and will improve program results only if the results are valid, transparent, and able to inform future program design and execution. In recent years, federal agencies have placed an increasing emphasis on demonstrating effectiveness through rigorous evaluations.[6] But there are concerns that funding levels have increased, while the efficiency and effectiveness of aid remain opaque and uncertain.[7]

---

4. U.K. DEP'T FOR INT'L DEV., ECONOMIC DEVELOPMENT STRATEGY: PROSPERITY, POVERTY AND MEETING GLOBAL CHALLENGES 27, 28 (2017).

5. GAO-17-316, *supra* note 3 (Executive Summary).

6. *Id.* at 7.

7. MARIAN LEONARDO LAWSON, CONG. RESEARCH SERV., R42827, DOES FOREIGN AID WORK? EFFORTS TO EVALUATE U.S. FOREIGN ASSISTANCE 1, 4 (2016), https://fas.org/sgp/crs/row/R42827.pdf.

B. Focus of FATAA

In response to these concerns, Congress passed the new FATAA legislation, which will impact all areas of U.S. foreign assistance.[8] The main focus of the legislation is a shift toward outcomes and impact of foreign assistance funding.[9] Agencies are required not only to measure outputs (i.e. number of kilometers built, malaria nets provided, or anti-corruption trainings held for judges), but also to assess outcomes and impacts (i.e. cost savings for vehicle owners, decreases in the prevalence of malaria, and corruption). This will ultimately improve program beneficiaries' incomes and standard of living.

C. New OMB and Agencies' Guidelines

FATAA serves as a catalyst for best practices in development monitoring and evaluation and has its framework set by the President of the United States in the form of OMB Guidelines to FATAA.[10] To meet the requirements outlined in FATAA, OMB issued guidelines on foreign aid evaluations in January 2018.[11]

The key development and security assistance agencies have also updated their monitoring and evaluation policies and guidance. Here are several examples:

- The U.S. Department of State (DOS) issued an integrated Program Design and Performance Management Toolkit in October 2016;[12] revised and updated a version of its program and project design, monitoring, and evaluation policy in November 2017;[13] and updated its guidance in 2018.[14] Subsequently, DOS issued an updated M&E

---

8. Office of Mgmt. & Budget, Exec. Office of the President, OMB Bull. No. 12-01, Guidance on Collection of U.S. Foreign Assistance Data 4 (2012) (defining Foreign assistance as tangible or intangible resources—goods, services, funds—provided by the USG to a foreign country or an international organization for the purpose of assistance to foreign entities or populations as authorized under the Foreign Assistance Act of 1961, as amended, or any other Act).

9. OMB Bull. No. 18-04, *supra* note 2, at 1.

10. *Id.* at 1.

11. *Id.* at 4.

12. U.S. Dep't of State, Program Design and Performance Management Toolkit (2016), https://www.state.gov/wp-content/uploads/2018/12/Program-Design-and-Performance-Management-Toolkit.pdf.

13. U.S. Dep't of State, Department of State Program and Project Design, Monitoring, and Evaluation Policy (2017), https://www.state.gov/wp-content/uploads/2018/12/Department-of-State-Program-and-Project-Design-Monitoring-and-Evaluation-Policy.pdf.

14. Bureau of Conflict and Stabilization Operations, U.S. Dep't of State, CSO Monitoring Guidance: Data to Inform, https://www.state.gov/wp-content/uploads/2018/12/CSO-Monitoring-Guidance-Data-to-Inform.pdf (last visited Nov. 3, 2019).

policy in January 2018 in compliance with the January 2018 OMB Guidelines.[15]

- The U.S. Agency for International Development (USAID) revised its Automated Directives System (ADS) Chapter 201 addressing evaluation guidance, planning, and implementation in September 2016 and later in October 2018.[16] USAID also developed toolkits to cover its work under FATAA—one for Monitoring, one for Evaluation, and one for Collaborating, Learning, and Adapting.[17]
- The Millennium Challenge Corporation (MCC) issued its March 2017 Policy for Monitoring and Evaluation,[18] which requires that compact M&E plans identify and describe its evaluation methodologies, key evaluation questions, and data collection strategies.
- The U.S. Department of Defense (DOD) issued agency-wide evaluation guidance for security cooperation in January 2017 in its DOD Instruction 5132.14 Assessment, Monitoring, and Evaluation Policy for the Security Cooperation Enterprise.[19]

III.   Current Condition of Foreign Assistance Evaluations' Quality, Cost, and Dissemination

A recent study conducted by the U.S. Government Accountability Office (GAO)[20] sheds some light on the current condition of M&E across USG.[21] GAO's study provides a baseline assessment of the quality, cost, and dissemination of foreign assistance evaluations when the legislation took effect.[22] It focuses on the six agencies providing "the largest amount of U.S.

---

15. *See* U.S. DEP'T OF STATE, *supra* 13.

16. U.S. AGENCY OF INT'L DEV., ADS CHAPTER 201 PROGRAM CYCLE OPERATIONAL POLICY 16, 50, 67 – 77 (2019), https://www.usaid.gov/ads/policy/200/201.

17. *See Monitoring, Evaluation, and CLA Toolkits*, LEARNING LAB, https://usaidlearninglab.org/mel-toolkits (last visited Nov. 3, 2019).

18. MILLENNIUM CHALLENGE CORP., POLICY FOR MONITORING AND EVALUATION 3 (2017), https://www.mcc.gov/resources/doc-pdf/policy-for-monitoring-and-evaluation.

19. U.S. DEP'T OF DEF., DoD INSTRUCTION 5132.14: ASSESSMENT, MONITORING, AND EVALUATION POLICY FOR SECURITY COOPERATION ENTERPRISE 3 (2017), https://open.defense.gov/portals/23/Documents/foreignasst/DoDI_513214_on_AM&E.pdf.

20. *See generally* GAO-17-316, *supra* note 3.

21. Other meta-evaluations conducted by the Lugar Center, Modernizing Foreign Assistance Network, and USAID have identified similar challenges and areas for improvement. *See generally* LUGAR CTR. AND MODERNIZING FOREIGN ASSISTANCE NETWORK, FROM EVIDENCE TO LEARNING: RECOMMENDATIONS TO IMPROVE U.S. FOREIGN ASSISTANCE EVALUATION (2017), http://www.thelugarcenter.org/assets/htmldocuments/TLC%20MFAN%20Evaluation%20Study%20Final%20112017.pdf; MGMT. SYS. INT'L, META-EVALUATION OF QUALITY AND COVERAGE OF USAID EVALUATIONS 2009-2012 (2013), https://www.usaid.gov/evaluation/meta-evaluation-quality-and-coverage; MGMT. SYS. INT'L, EVALUATION UTILIZATION AT USAID (2016), https://www.usaid.gov/documents/1870/evaluation-utilization-usaid.

22. GAO-17-316, *supra* note 3, at 2.

foreign assistance:" USAID, DOS, MCC, USDA, HHS, and DOD.[23] The study found that about three quarters of the 170 evaluations completed in fiscal year 2015 by these agencies and reviewed by GAO generally or partially addressed all of the quality criteria GAO identified for evaluation design, implementation, and conclusions.[24]

Agencies met some elements of the GAO quality criteria more often than other elements. For example, approximately ninety percent of all evaluations addressed questions that are generally aligned with program goals and were thus able to provide useful information about program results.[25] In addition, about eighty percent of evaluations had an appropriate design given the study questions.[26] Moreover, "indicators for measuring progress were generally appropriate in about eighty percent of the evaluations."[27] But approximately forty percent of evaluations had various deficiencies in implementing their intended designs.[28]

A. EVALUATION DESIGN

A common challenge in evaluation design was the attempt to answer impact questions with a *qualitative* design based on interviews with a non-representative sample of participants. For example, questions in some evaluations were phrased in terms of effectiveness and changes over time without any clearly stated counterfactual.[29] In addition, while almost all evaluations attempted to assess processes and outcomes using output and outcome metrics, approximately half did not have baselines, specific targets, or criteria against which to assess progress.[30] For example, an evaluation of a literacy program focused on certification rates for those present at the end of the program, but there were no pre-enrollment measures, and the attrition rate was known to be high.[31]

Most evaluations used appropriate indicators and measures for the study objectives.[32] But some evaluations used questions that asked respondents to identify how much they attribute the positive changes to the program, rather than using indicators of change in attitudes based on pre– and post–measures of change. For example, an evaluation of a training program used "satisfaction surveys," rather than "knowledge surveys," to assess the results of the training. An evaluation in the same sector asked respondents

---

23. *Id.* at 1.
24. *Id.* at 8, 11.
25. *Id.* at 13.
26. *Id.* at 14.
27. *Id.*
28. *Id.*
29. *See* LUGAR CTR. AND MODERNIZING FOREIGN ASSISTANCE NETWORK, *supra* note 21, at 25.
30. GAO-17-316, *supra* note 3, at 16.
31. *Id.* at 12.
32. *See* LUGAR CTR. AND MODERNIZING FOREIGN ASSISTANCE NETWORK, supra note 21, at 36.

for "input from the interviewees" as an indicator of program success in order to judge the extent to which they have learned how to read and write. Another evaluation included a list of indicators, but those indicators were not subsequently used to assess the success of the intervention. Similarly, one evaluation included a list of indicators, but there was no clear link between the indicators and the logical framework of the program. Often evaluations included no indicators and instead included broad statements of progress or behavioral outcome changes.

B.   EVALUATION IMPLEMENTATION

A key element in assessing evaluation implementation is the extent to which target population and sampling, data collection, and data analysis were appropriate for the study questions. Because the target population is the group the researcher would like to make statements in the evaluation, it is important that the group is clearly defined and includes all potential beneficiaries of the program. In addition, both random and non-random sampling from the target population could be and were used in evaluation implementation. That said, *random* sampling was appropriately implemented in most evaluations, but *non-random* sampling was not appropriately implemented in approximately half of the evaluations. Some evaluations included in the target population only individuals who agreed to participate in the program but did not take into account those who declined. Other studies identified the target population with broad statements such as "all beneficiaries" or "several types of stakeholders" or descriptions of age and location, but no information of its size.[33] For some evaluations, complete, accurate, and up-to-date lists of program participants were not available.[34] For others, it was hard to understand how the lists of contacts and participants were developed. In others, the method of participant selection was unclear, especially for the collection of testimonial evidence.[35] In one evaluation the evidence was collected from less than twenty program office staff members who were asked to describe the effectiveness of the program and select program participants for subsequent interviews based on recommendations from the program office. Thus, selection bias—in particular, self-selection—was a significant concern as the selection process, and criteria were not clearly defined.

Evaluations used a variety of *data collection* methods. Most evaluations used interviews, program data, and focus groups, while about forty percent used surveys and direct observation techniques.[36] Some common areas of concern included unclear quality control processes for data collection in

33. *See, e.g.*, Cornelia Loechl et al., *An Operations Evaluation of World Vision's Maternal and Child Health and Nutrition Program in Central Plateau, Haiti*, FOOD AND NUTRITION TECHNICAL ASSISTANCE (FANTA) PROJECT 10, 27 (2004), https://pdfs.semanticscholar.org/5e8e/66cf2a23c2f09371ba75bc31c0e407b9c7dd.pdf.
34. GAO-17-316, *supra* note 3, at 15.
35. *Id.*
36. *Id.* at 15 n.24.

terms of who collected the data and how the data collection instruments were developed, tested, and administered.[37] In some cases, a bias toward self-reporting was present. For example, a problematic case involved mentors who were asked to collect data and report on how effective they thought they were. Overall the data collection procedures generally appeared to ensure the reliability of the data in less than half of the cases, with missing data and insufficient sample size for the findings as some of the main concerns.

Evaluations used a mix of *qualitative* and *quantitative* data analysis techniques, which were assessed as appropriate in the majority of the evaluations. Some potential areas for improvement included a more explicit description of the steps taken for content analysis of the qualitative evidence. For example, in some cases it was not clear how the interviews were analyzed and how the categories for content analysis were selected.[38] In another case, reporting of data did not identify the source or extent of the evidence supporting the statements, and there was no discussion of how the summaries of the interviews were performed. For evaluations that used focus groups, some did not specify how they accounted for the influence of focus group participants on each other.[39] In addition, the lack of quantitative information made it difficult in some cases to answer questions addressing tangible outcomes and project achievements. Some evaluations that used quantitative data analysis methods did not describe the performed statistical tests, robustness checks, and sensitivity analysis and did not state the assumptions for the statistical analysis.

C.   Evaluation Conclusions

Despite the various limitations in the design and implementation of the evaluations, overall conclusions generally considered the strengths and limitations of the evidence and did not overreach in their statements. About seventy percent of the evaluations contained conclusions that were carefully worded and well supported by the evidence, and the recommendations and lessons learned were found to be justified.[40] But some conclusions went beyond the collected evidence and did not appropriately reflect the gathered information. For example, one evaluation did not specify what effect the lack of key informants had on the evaluation conclusions and how informants on "approved lists" affected the results. In another evaluation, the conclusions were linked to the findings, but the findings were not based on identifiable evidence. Similarly, another evaluation concluded that the intervention was effective, but the evaluation focused on implementation issues.

---

37. *Id.* at 15.
38. *Id.* at 16.
39. *Id.*
40. *See* GAO-17-316, *supra* note 3, at 17.

D. AGENCY-SPECIFIC RESULTS

Implementing evaluations overseas poses significant methodological challenges, and the GAO study identified opportunities for each agency to improve evaluation quality and thereby strengthen its ability to manage aid funds more effectively based on results. About a third of evaluations by USAID, MCC, and HHS were assessed as high-quality evaluations, while close to half of DOS and DOD evaluations were placed in the lower-quality category.[41] In addition, each criterion was met by the majority of USAID, MCC, and HHS evaluations, but they generally scored lower in evaluation implementation.[42] DOD guidance for the Global Train and Equip Program requires that a baseline assessment of recipient unit capabilities be completed prior to submission of the project proposal. But some of the four assessment reports GAO reviewed for the Global Train and Equip Program had a number of shortcomings. For example, they did not discuss who the target population was, how the equipment was selected, how the respondents were chosen for interviews, what the baselines and targets were, how the evidence was analyzed, and what effect any possible limitations might have had on the stated conclusions of the studies.

E. EVALUATION COSTS AND DISSEMINATION

The March 2017 GAO study opens its report with "What GAO Found," an executive summary of the report, which states as follows:

Evaluation costs ranged widely and were sometimes difficult to determine, but the majority of evaluations GAO examined cost less than $200,000. [MCC's] evaluations had a median cost of about $269,000, while median costs for the [USAID, USDA, and DOS] ranged from about $88,000 to about $178,000. GAO was unable to identify the specific costs for the [DOD and HHS] evaluations.[43]

A higher quality of evaluation tended to correlate with a higher cost of conducting the evaluation, but "some well-designed[,] lower-cost evaluations also met all quality criteria."[44] Other factors relating to evaluation costs included the evaluation's data-collection methods and frequency of data collection, as well as the evaluation's duration and design. Evaluations that included primary data collection cost more than those which did not. In particular, evaluations with a survey of beneficiaries typically cost about $74,000 more than those without a survey.[45] Evaluations which included data collection repeated over time cost about $44,500 more than those which did not.[46] The use of randomized controlled trials and quasi-experimental

---

41. GAO-17-316, *supra* note 3, at 11.
42. *Id.*
43. GAO-17-316, *supra* note 3 (Executive Summary).
44. *Id.*
45. *Id.* at 23 – 24.
46. *Id.* at 23.

designs also increased evaluation costs.[47] For example, "five of the six evaluations that cost more than $900,000 were net impact evaluations."[48] Although agencies "generally posted and distributed evaluations for the use of internal and external stakeholders," the "shortfalls" in public posting, timeliness, and dissemination planning may place limits on the evaluations' usefulness.[49]

IV.   Challenges in Implementing the New Development Framework

The international development community is waiting to see what is feasible in terms of attaining the best practices put forth in the FATAA, OMB guidelines, and updated agencies' policies and guidelines. While the key agencies have already put in place new guidelines and standards around M&E as well as data policies, funders and implementers may need to become more flexible to enable foreign aid to accomplish its mission in the best interests of national security and other goals of U.S. foreign assistance. Some factors to consider include the size of projects and implementers, the timeline and nature of certain reforms, and, ultimately, the strategic objective set forth at the design stage of each intervention.

A.   SIZE OF PROJECTS AND IMPLEMENTERS

Foreign assistance programs, projects, and activities vary in dollar value and are implemented by a diverse set of partners, which varies in size and other characteristics. For example, small implementers and large implementers are given the same mandate of best practices to follow, yet the smaller implementers do not necessarily have the means that a larger implementation project would allow. For example, a $1 million, two-year project for the USPTO to train judges in Pakistan on the enforcement of intellectual property may not have the same means to conduct a rigorous evaluation as a $50 million, five-year project to a single private investor to train judges in intellectual property rights in Sri Lanka. There can be several implementers in a State Department five-year, $200 million project to improve intellectual property rights enforcement in South Asia. There are more questions raised by the new law and guidelines for implementers: when is an outside evaluator justified and when should an implementer conduct evaluations in-house? The new guidelines increase focus on evaluation, but more guidance is necessary on all levels of funding agencies to ensure that resources are made available for studies with appropriate methodologies to identify outcomes and impact of the interventions.[50]

---

47. GAO-17-316, *supra* note 3, at 21.
48. *Id.* at 21.
49. *Id.*
50. For a comprehensive review of evaluation methods, see JUDY L. BAKER, EVALUATING THE IMPACT OF DEVELOPMENT PROJECTS ON POVERTY: A HANDBOOK FOR PRACTITIONERS (2000).

B. LONG-TERM EFFECTS AND NATURE OF CERTAIN INTERVENTIONS

   Methods for measuring effectiveness and impact of various interventions in a particular area must be adapted to the type of programming implemented, sensitivities, geographic reach, and political context. Critics of the overly bureaucratic processes around foreign aid measurement argue that some of the least important projects are the easiest to evaluate and the most transformational projects are the least measurable.[51] Therefore, the focus on "bean counting" may lead to overlooking what is truly important in terms of development outcomes.[52] Certain outcomes take a longer time to emerge and may be less tangible. The Department for International Development (DFID) of the United Kingdom, for example, recognizes that value for money of policy work may be harder to track and quantify.[53] Nonetheless, governments increasingly commit to go beyond mechanically tracking the outputs and collect data that would pave way for further analysis. Interventions in certain sectors, such as Democracy, Rights, and Governance (DRG) or Rule of Law Development, are long-term in nature and need a careful balance of methods to document, track, and demonstrate changes over time in these complex settings. In recent years, USG has reformed its approach to focus on outcomes and impact measurement, and the FATAA is an effort to codify this approach. But much remains to be done.

   Scholars and practitioners acknowledge the long-term nature of successful development programming in the area of rule of law, among others, and experts place continued emphasis on sustainability for at least fifty years.[54] Debate persists[55] around what types of methods for measuring success are appropriate in complex environments such as government effectiveness.[56] Traditionally, randomized controlled trials (RCTs) were widely applied in evaluations of projects in health, agriculture, economic development, and education; however, they were less applied in other sectors, such as Democracy and Governance (DG).[57] In the last decade, there have been several studies funded by U.S. agencies to assess impact through RCTs on

---

   51. ANDREW NATSIOS, THE CLASH OF THE COUNTER-BUREAUCRACY AND DEVELOPMENT 3 (Ctr. For Global Dev. 2010).

   52. *Id.*

   53. U.K. DEP'T FOR INT'L DEV., *supra* note 4, at 30.

   54. RACHEL KLEINFELD, IMPROVING DEVELOPMENT AID DESIGN AND EVALUATION: PLAN FOR SAILBOATS, NOT TRAINS 4, 37 (2015).

   55. Martin Ravallion, *Should the Randomistas (Continue to) Rule?* 3 (Ctr. for Global Dev., Working Paper No. 492, 2018).

   56. Rachel M. Gisselquist & Miguel Niño-Zarazúa, *What Can Experiments Tell Us About How to Improve Government Performance?*, 6 J. OF GLOBALIZATION AND DEV. 1, 4 – 5 (2015).

   57. JACK A. GOLDSTONE ET AL., NAT'L RESEARCH COUNCIL, COMM. ON EVALUATION OF USAID DEMOCRACY ASSISTANCE PROGRAMS, IMPROVING DEMOCRACY ASSISTANCE: BUILDING KNOWLEDGE THROUGH EVALUATIONS AND RESEARCH 145 (2008).

elements of programming in the DG sector.[58] Arguably a mixture of methods and indicators, from counting outputs to observing outcomes and measuring impact, would be optimal depending on the context, availability of data, and objectives of aid.[59] Implementers and funders should consider interdisciplinary research as a suitable approach to studying outcomes in complex areas such as rule of law.[60] This type of research is known to take more time and effort.[61] But it may yield more insight into the contextual factors and provide theoretical foundations necessary to understand the complexity of cultural, social, and legal norms and practices that are often involved in development outcomes. As difficult as it may be,[62] policymakers need to account for the timelines related to ambitious change agendas, consider long-term engagement in countries, and explore a variety of methods to understand long-term changes as a result of foreign assistance implementation.

C.   THE CASE OF THE AMERICAN BAR ASSOCIATION'S RULE OF LAW INITIATIVE

One of the implementers in the rule of law development space is the American Bar Association through its Rule of Law Initiative (ABA ROLI). More than twenty-five years ago, the ABA started its work abroad through the Central European and Eurasian Law Initiative (CEELI).[63] CEELI provided technical assistance to former Soviet states after the fall of the Soviet Union with the goal of creating democratic reforms in those countries.[64] As the ABA expanded CEELI's work past Europe to accommodate a now global programming that spans across Africa, Asia, Latin America, the Middle East, Caribbean, and Northern Africa, CEELI was brought under a Rule of Law Initiative (ROLI) in 2008.[65] Today, ABA ROLI does work in more than fifty countries promoting rule of law through various types of programming, from creating electronic access to courts to bar strengthening.[66]

---

58. *See generally List of DRG Funded Research*, U.S. AGENCY FOR INT'L DEV., https://www.usaid.gov/node/33416 (last visited Nov. 3, 2019).

59. M.G. Quibria, *Aid Effectiveness: Research, Policy and Unresolved Issues*, 1 DEV. STUD. RES. 75, 83 (2014).

60. KRISTINA SIMION, QUALITATIVE AND QUANTITATIVE APPROACHES TO RULE OF LAW RESEARCH 25 (2016).

61. *Id.*

62. MARIAN LEONARDO LAWSON, CONG. RESEARCH SERV., R42827, DOES FOREIGN AID WORK? EFFORTS TO EVALUATE U.S. FOREIGN ASSISTANCE 1, 14 (2016), https://fas.org/sgp/crs/row/R42827.pdf.

63. M. Margaret McKeown, *The ABA Rule of Law Initiative Celebrating 25 Years of Global Initiatives*, 39 MICH. J. INT'L L. 117, 121, 126 – 27 (2018).

64. *Id.* at 121.

65. *Id.* at 128.

66. AM. BAR ASS'N RULE OF L. INITIATIVE, ANNUAL REPORT 2016-2017, at 3, 4, 10, 17, 20 (2017).

This is a complex and extensive endeavor, and ABA ROLI has been striving to assess its effectiveness and impact on long-term outcomes. Some of the examples of long-term programming include facilitating the establishment of bars or bar associations, implementing reforms around continued legal education for practicing lawyers, and supporting curriculum development for law schools.[67] In Kazakhstan and Kosovo, ABA ROLI aided the drafting of legislation relevant to the legal profession.[68] In Georgia, it provided support for round tables and the Advocates Association's promulgation of amendments to the Law of Advocates.[69] In Kyrgyzstan, ABA ROLI supported advocacy around the Law on Advocatura (the criminal defense bar) for over ten years and saw it come into force in 2015.[70] The latter program was evaluated internally using an outcome harvesting approach to establish organizational contribution to the engagement for over a decade.

In Armenia, ABA ROLI has been improving access to justice by strengthening the legal profession and aiding in the creation and strengthening of a Public Defender's Office (PDO).[71] Since its founding in 2006, the PDO has provided legal representation to Armenian citizens based on statutorily-defined eligibility criteria in criminal and civil cases.[72] In the years since its establishment, the PDO staff has increased its capacity to process cases.[73] In a recent case study of Armenia, where the judiciary and legal clinics work to improve Armenians' access to the European Court of Human Rights (ECHR), researchers found citizens experiencing rights violations turned to lawyers and human rights organizations for legal help. They are viewed as respectable and reliable resources for remedy. This is a significant shift because the public previously considered them part of a corrupt legal system. By 2016, the ABA ROLI-supported Strategic Litigation Fund had filed six cases to the ECHR. While multiple factors may influence such shifts, including, for example, Armenia joining the Council of Europe—which made the ECHR mechanism available to the public—ABA ROLI's long-term presence in countries has enabled it to identify certain outcomes and establish plausible contribution using a combination of methods. From the perspective of a Rule of Law

---

67. *Id.* at 4.

68. *Id.* at 10, 17.

69. *Id.* at 10.

70. *See National Association of Criminal Defense Lawyers Conducts its Founding Congress*, AM. BAR ASS'N: RULE OF L. INITIATIVE (Apr. 1, 2015), https://www.americanbar.org/advocacy/rule_of_law/where_we_work/europe_eurasia/kyrgyzstan/news/news-kyrgyzstan-advocatura-conducts-founding-congress-0415/.

71. *Armenia Background*, AM. BAR ASS'N: RULE OF L. INITIATIVE, https://www.americanbar.org/advocacy/rule_of_law/where_we_work/europe_eurasia/armenia/background.html (last visited Nov. 3, 2019).

72. McKeown, *supra* note 63, at 135 – 36.

73. THE EUROPEAN COMM'N FOR THE EFFICIENCY OF JUSTICE, COUNCIL OF EUROPE, 1. EVALUATION OF THE JUDICIAL SYSTEMS (2016-2018 CYCLE): ARMENIA 10 (2018), https://rm.coe.int/armenia/16808d0248.

Development implementer, several methods have proven effective in measuring outcomes in rule of law, including contribution analysis, Outcome Harvesting, and Outcome Mapping. Contribution analysis[74] is a method that helps address a key question for development programs: one of attribution. While experimental and quasi-experimental approaches may not always be feasible, contribution analysis can help implementers come to robust conclusions whether the intervention added value and how it contributed to the observed results.[75] Outcome Harvesting[76] and Outcome Mapping,[77] as the names suggest, focus on understanding outcomes that emerge from the initiative's activities. But these approaches look further into the future and are broader in scope in order to assess longer-term economic, environmental, political, or demographic changes or societal shifts.

D.   INTERAGENCY COORDINATION

A significant element of foreign assistance implementation is interagency cooperation. Foreign assistance funding implementers throughout the USG provide foreign assistance programming in conjunction with each other. Often, foreign assistance funds earmarked for a particular agency, such as USAID, are transferred from the originating or *funder* agency to another agency or *recipient*. Similar to funds transferred into the private sector via grants, funding agencies conduct interagency transfer through 632(a) or 632(b) Inter-Agency Agreements (IAAs).[78] A typical IAA would be from the DOS to the DOJ in order to fund the DOJ's anti-corruption training for judges in a developing country. See Table 1 (below) for a breakdown of foreign assistance transfer funds by DOS and USAID by implementing agency and funding account for fiscal year 2017, the latest publicly available complete dataset.

Table 1: Transfer Funding From the Department of State and USAID to Implementing Agencies By Funding Account, Fiscal Year 2017

Funding Agency–Department of State	$ 27,653,975,621
Implementing Agencies (Other Than Department of State):	
Department of Defense	$ 310,599,179

---

74. John Mayne, *Contribution Analysis: An Approach to Exploring Cause and Effect*, ILAC BRIEF 16 1 (2008), https://cgspace.cgiar.org/bitstream/handle/10568/70124/ILAC_Brief16_Contribution_Analysis.pdf?sequence=1&isAllowed=y.

75. *Id.*

76. *See generally* RICARDO WILSON-GRAU, OUTCOME HARVESTING: PRINCIPLES, STEPS, AND EVALUATION APPLICATIONS (Jennifer C. Greene ed., 2018).

77. *See generally* SARAH EARL ET AL., OUTCOME MAPPING: BUILDING LEARNING AND REFLECTION INTO DEVELOPMENT PROGRAMS (2001).

78. *See* 22 U.S.C. § 2392 (2019).

Department of State, Global Health Programs	$ 308,196,916
Department of State, Global HIV/AIDs Initiative	$ 2,402,263
Department of Health and Human Services	$ 4,050,050,571
Department of State, Global Health Programs	$ 4,001,253,977
Department of State, Global HIV/AIDs Initiative	$ 48,796,594
Department of Homeland Security	$ 8,095,630
Nonproliferation, Anti-Terrorism, Demining and Related Programs	$ 8,075,323
Peace Keeping Operations	$ 20,307
Department of Justice	$ 3,931,467
International Narcotics Control and Law Enforcement	$ 3,207,110
Nonproliferation, Anti-Terrorism, Demining and Related Programs	$ 724,357
Funding Agency-U.S. Agency for International Development	$ 27,545,330,764
Implementing Agencies (Other Than USAID):	
Department of Commerce	$ 8,054,716
Economic Support Fund	$ 8,054,716
Department of Homeland Security	$ 29,630
Agency for International Development, Development Assistance	$ 29,630
Department of State	$ 1,927,713,969
Agency for International Development, Development Assistance	$ 497,919
Assistance for Europe, Eurasia and Central Asia (AEECA)	$ 194,850,133
Economic Support Fund	$ 1,732,365,917
Department of the Interior	$ 3,935,130
Agency for International Development, Development Assistance	$ 3,935,130
Environmental Protection Agency	$ 8,237,341
Agency for International Development, Development Assistance	$ 540,548
Economic Support Fund	$ 7,696,793
Inter-American Foundation	$ 4,662,053
Agency for International Development, Development Assistance	$ 4,662,053
Trade and Development Agency	$ 14,964,127
Economic Support Fund	$ 14,964,127

Source: Authors' analysis of USAID Foreign Aid Explorer Data.

According to OMB M&E Guidelines for foreign assistance, all federal departments and agencies that administer U.S. foreign assistance must develop

specific policies and procedures for the M&E of their programs.[79] These policies must address funding transfers between or among USG agencies and ensure accountability for M&E, including in cases where one agency leads or coordinates an overall program but multiple agencies implement activities under that program. Furthermore, the guidelines require that M&E roles and responsibilities be (1) considered and documented when funds are transferred; (2) defined when other agencies will be implementing activities that support a multi-agency project for a lead agency; (3) clearly defined in interagency agreements; and (4) shared along with necessary assessments, past evaluations, and other information by lead agencies to assist supporting agencies with their M&E.

Similar to evaluation for foreign assistance programming conducted by a single agency, evaluation for programming provided by multiple collaborating agencies must evaluate the outputs and outcomes of the joint assistance programming. While the guidelines clearly require a delineation of funding and implementing agencies' roles and responsibilities with regard to M&E, various aspects of interagency coordination require further clarification.

1. Interagency Program Definition

The definition of a "program" in FATAA implementation could substantially impact agencies that receive foreign assistance funds through IAAs. For example, FATAA requires impact analysis for programs of a representative size, which can be cost-prohibitive for relatively small agencies. As a result, some recipient agencies that conduct foreign assistance activities may not have funds to conduct M&E in addition to the M&E required for the broader and larger "program" that includes all agencies involved. Therefore, a "program" should be defined as the broader and larger effort by all agencies involved in reaching a particular goal in a country or region, not the specific activities requested in an IAA. Individual agencies that receive funds through IAAs should not be expected to conduct M&E of programs outside of that broader program definition. But we note that the higher the level of aggregation of the unit of evaluation in terms of an activity, project, or program, the more difficult it is to conduct a rigorous impact assessment. Therefore, rigorous impact assessments at the program-level should be designed early—at the program's inception—because a valid counterfactual may be difficult to establish later on and many confounding factors may influence the overall impact of the program.

2. Collaboration on Logic Model Development

During the IAA development process, agencies that *provide* foreign assistance funds through IAAs generally identify indicators for the agencies that receive funds through IAAs to measure the particular goals and objectives identified for the programs. Agencies that receive funds through

---

79. OMB BULL. NO. 18-04, *supra* note 2 (Introductory Statement).

IAAs should be empowered to coordinate indicators with funders identified through a *consensus* of the IAA funder and IAA recipient. For example, the DOS may fund the DOJ to provide demand-side anti-corruption training for judges in Colombia and may fund the U.S. Department of Commerce (DOC) to provide supply-side anti-corruption programming for private companies in Colombia. The outcome of this training may be to improve Colombia's Corruption Perception Index (CPI) as measured by Transparency International. While the efforts of the individual agencies may be tracked using appropriate output indicators, both DOJ and DOC can point to Colombia's CPI as an outcome of their programming. In this case, the DOS would need to integrate the activities of individual agencies and coordinate logic model development of both agencies to determine that CPI is the intended overall outcome measurement of the joint efforts.

In addition, as logic models and counterfactual narratives are important for determining success, failure, and course corrections, the agencies that receive funds through IAAs will be dependent upon the logic models and counterfactuals determined by its funders for the whole program. Agencies that receive funds through IAAs should have the opportunity to provide context to the defined inputs, activities, outputs, and outcomes expected to be measured for M&E. Also, agencies that receive funds through IAAs should have the opportunity to provide context to the counterfactual narrative in identifying program effect.

3. *Attribution of Outcomes to Individual Agencies*

Interagency collaboration requires that funders determine foreign assistance program planning and evaluation in conjunction with recipient agencies, but attribution of outcomes to individual agency efforts may be problematic. For example, in the case involving anti-corruption programming in Columbia, DOS funds several agencies to achieve the goals of the overall program. Therefore, DOS may be coordinating an evaluation of several agencies that are trying to reach the same outcome in a particular jurisdiction, but attribution of that outcome to a particular agency's efforts will be unlikely. This reinforces the necessity of M&E coordination between funder and recipient agencies throughout the lifecycle of a program—from planning to final reporting.

4. *Coordinated M&E Frameworks in Joint Program Implementation*

Agencies that receive funds through IAAs may receive funds from many different sources, including DOS and USAID geographical and functional bureaus, as well as U.S. Embassies or USAID Missions abroad. Recipient agencies can have several different funding sources in over dozens of IAAs. Agencies that receive funds through IAAs may be asked to maintain many different M&E systems from the many funding sources and IAAs that codify the transfers. Therefore, agencies that receive funding from IAAs should be allowed to propose consistency among M&E systems designed by the

agencies funding them. For example, if USAID and DOS had standard M&E systems that would apply to all IAA transfers, this would reduce the burden on agencies that receive funding from multiple sources.

5. OMB Reporting

The obvious need for better definition of roles and responsibilities of funding and implementing agencies of U.S. foreign assistance programs and their evaluation has spawned interagency policy committee meetings in the USG. In these meetings, OMB has advised interagency foreign assistance funders and implementers regarding foreign assistance evaluation and has established standard policy for foreign assistance interagency cooperation. Recipient agencies implementing programs, projects, or events with funds from funding agencies will not necessarily report directly to OMB on those programs, projects, or events, but rather report through their funders. OMB will follow up with the funding agencies during budget negotiations regarding progress on FATAA implementation. For example, if an agency's funds come from USAID, then USAID will be at the table with OMB during those budget discussions. As agencies that receive foreign assistance funds through IAAs may not meet with OMB directly during the budget process, there is required coordination between funding and recipient agencies at all points of IAA development and implementation.

6. Additional Considerations

There are a number of open questions and potential implications of enhanced evaluation requirements stemming from the new FATAA legislation, OMB guidelines, and updated agencies' guidance. A few of those questions include:

- High quality evaluations may require substantial resources. How are those costs estimated and who pays for them?
- Use of evaluation results requires publication and transparency. What are the best practices to disseminate evaluation findings and encourage management to incorporate lessons learned in the design of future projects?
- Pressure to evaluate may lead to a redistribution of resources from programming to evaluation. Are we ready to shift funding to a new class of evaluators and make less aid available to the would-be beneficiaries?
- Foreign aid implementers vary in size, skill sets, and capacity for evaluation. Do the requirements unfairly burden small implementers and those they serve, taking away limited time and resources from their development aid mission?
- Rigorous evaluation requirements may favor centralized management of foreign aid projects as opposed to management in the field. What are the implications of such a shift for program management and implementation?

V. Conclusion

The FATAA has significant implications for the evaluation and learning agendas of USG agencies and across USG-funded implementers. The focus by USG agencies on outcomes is a positive development that strengthens results-based management, as monitoring alone is no longer sufficient. We need to invest in conducting strong baseline assessments so that eventually outcomes and impact can be traced. Decisionmakers on all levels need to consider the variety of foreign assistance interventions and the many ways in which measurement will work for each of them. We need to also increase collaboration across various USG agencies, speak consistently to the implementers, and build on each other's work on baseline and evaluation measurements.

Yet, the shifted focus on more rigorous evaluation has raised some concerns about potential trade-offs and the universal applicability of the new requirements. For example, not everything in development assistance is readily measurable, knowable, or provable through empirical study or measurable at a reasonable cost. Moreover, some worthwhile programs take a long time to measure or do not progress in linear fashion, as they show progress in a circuitous route (like a sailboat rather than a train trajectory).[80] Therefore, perhaps application of the guidelines and standards need to be flexible in light of these considerations to enable foreign aid to accomplish its mission in the best interests of the stated goals of U.S. foreign assistance.

---

80. *See* KLEINFELD, *supra* note 54, at 37; *see also* Norman L. Greene, *Monitoring and Evaluation of International Counter-Trafficking Programs: Definitions, Challenges, and a Way Forward,* 5 ST. JOHNS J. OF INT'L AND COMPARATIVE L. 153, 165 (2015).