USAID
FROM THE AMERICAN PEOPLE

BENOIT ALMERAS/ HANDICAP INTERNATIONAL

DRG IMPACT EVALUATION RETROSPECTIVE:

Learning from Three Generations of Impact Evaluations

A Learning, Evaluation, and Research Activity II (LER II)

October 2022 (Originally published April 2021)

This DRG Impact Evaluation (IE) Retrospective was conducted by Dr. Michael G. Findley of the University of Texas at Austin; Ms. Aleta Starosta, Evaluation Specialist at The Cloudburst Group; and Dr. Daniel Sabet of USAID's Democracy, Human Rights and Governance (DRG) Center, with support from Ryan Hatano of The Cloudburst Group. The evaluation team would like to thank the USAID staff, implementing partners, principal investigators, and evaluators who shared their experiences and expertise with us.

This document was produced for review by the United States Agency for International Development, Democracy, Human Rights, and Governance Center under the Learning, Evaluation, and Research Activity II (LER II) contract: GS10F0218U/7200AA18M00017.

Prepared by:

The Cloudburst Group
8400 Corporate Drive, Suite 550
Landover, MD 20785-2238
Tel: 301-918-4400

CONTENTS

ACRONYMS

ADS	Automated Directives System
COI	Conflict of Interest
COR	Contracting Officer Representative
CSO	Civil Society Organization
C-TIP	Countering Trafficking in Persons
DEC	Development Experience Clearinghouse
DRC	Democratic Republic of the Congo
DRG	Democracy, Human Rights, and Governance
EDGE	Evaluating Democracy and Governance Effectiveness
GAO	Government Accountability Office
GAPP	Governance, Accountability, Participation, and Performance
GSAM	Ghana Strengthening Accountability Mechanisms
IE	Impact Evaluations
ILAB	Bureau of International Labor Affairs
IP	Implementing Partner
IPA	Innovations for Poverty Action
J-PAL	Abdul Latif Jameel Poverty Action Lab
KII	Key Informant Interview
LER	Learning, Evaluation, and Research
LGAP	Local Governance Accountability and Performance
LP	Learning Partner
M&E	Monitoring and Evaluation
MCC	Millennium Challenge Corporation
MEL	Monitoring, Evaluation, and Learning
MOU	Memorandum of Understanding
NDI	National Democratic Institute
PE	Performance Evaluation
PI	Principal Investigator
PPL	Bureau for Policy, Planning, and Learning
RCT	Randomized Control Trial
RFP	Request for Proposal
SMS	Short Message Service
USAID	United States Agency for International Development

EXECUTIVE SUMMARY

INTRODUCTION

In response to an influential 2008 National Academies of Sciences report, the United States Agency for International Development's (USAID's) Democracy, Human Rights, and Governance (DRG) Center initiated a pilot program of impact evaluations (IEs).[1] As of March 2021, since 2012, the DRG Center and its learning partners (LPs) have completed or are close to completing 27 IEs. This retrospective intends to provide a look back at both the accomplishments and the challenges of DRG Center IEs with the goal of deriving lessons learned and providing evidence-based recommendations for future DRG Center evaluation work.

RETROSPECTIVE QUESTIONS

This retrospective answers the following five questions:

1. **Description**: How many IEs were initiated, and how many were completed? What was the cost of these evaluations, and what topics and regions did they target? What methodologies were used? For those that were not completed, why were they not completed?
2. **Findings**: At a high level, what has USAID learned from the findings of these IEs?
3. **Challenges and lessons learned:** What have been the challenges encountered in designing and carrying out IEs, and what are the related lessons learned (for the DRG Center, Missions, implementing partners (IPs), and evaluators)?
4. **Use:** How has USAID (or others) used the IEs? Why were some evaluations more useful than others? How have findings been disseminated?
5. **Recommendations**: What should be the DRG Center's approach to IEs moving forward? Under what conditions are they most effective and useful? How could the DRG Center better support Missions and others in the utilization of IE findings/recommendations?

METHODOLOGY

This retrospective relies on a mixed-methods design including a combination of desk-based research, individual and group key informant interviews (KIIs), and an online survey offered to stakeholders in all previous DRG Center IEs. Although the desk review and survey targeted all 27 IEs, the evaluation team purposively selected eight IEs based on the strength of the theory of change, implementation challenges, findings, and use for more in-depth interviews with evaluation stakeholders. These included representatives from USAID, IPs, evaluators, and principal investigators (PIs). Of the 127 individuals invited to respond to the survey, 80 participated, yielding a response rate of 63 percent, and 64 individuals participated in KIIs.

[1] Goldstone, Jack A., Larry Garber, John Gerring, Clark C. Gibson, Mitchell A. Seligson, Jeremy Weinstein (2008) *Improving Democracy Assistance: Building Knowledge through Evaluations and Research.* Washington DC: National Academies Press.

DESCRIPTION OF DRG CENTER IMPACT EVALUATIONS

The DRG Center's approach incorporated several lessons learned from an initial generation of IEs to build a successful IE program. It included a flexible contracting mechanism, the involvement of top academics, and strategies to build Mission buy-in, such as training, multi-day IE workshops between academics and Mission staff (known as IE Clinics), and co-funding. The 27 completed or close-to-complete IEs covered a range of DRG issues and were geographically dispersed. Nine planned IEs did not move past the design stage for a variety of reasons, and two IEs were cancelled after baseline data collection. Of those for which the team has data, the median IE cost was $557,582 and the average cost was $713,202, which is on par with other USAID offices and other IE contracting organizations. In many ways, the DRG Center's IE program was a model in creating academic-Mission linkages to implement a robust IE program.

LEARNINGS FROM THE BENEFITS OF IMPACT EVALUATIONS

Unlike more traditional performance evaluations (PEs) and monitoring, IEs are able to measure a counterfactual for an intervention and make causal inferences about that activity's impact. In Haiti, an IE demonstrated that a program was working and should be scaled up. In the Caribbean, an IE found that previous project reviews, which concluded that the intervention was producing dramatic results, were incorrect. Furthermore, IEs frequently provided better measures of outcomes and changes in those outcomes over time relative to the earlier studies that did not employ counterfactual reasoning rigorously as per USAID IE guidelines. In some cases, baseline data or regression analysis produced insightful information that implementers could use to shape their programming. As such, the evaluation team can point to valuable findings that would not have otherwise existed in the absence of an IE.

Notably, although IEs encountered a number of challenges, the vast majority of stakeholders interviewed and surveyed acknowledged the indispensable role of impact evaluation and they expressed their support for the continuation of impact evaluation within DRG and USAID. As a general point about this retrospective, on balance, stakeholders are overwhelmingly positive about the role of impact evaluation, they encourage its continuation, and their negative commentary is offered in the spirit of constructive criticism designed to improve a fundamentally well-intentioned, but not fully developed, model. With that in mind, we note some of the key areas stakeholders hoped to improve.

LESSONS LEARNED FROM IMPACT EVALUATION CHALLENGES

Despite significant achievements, the DRG Center's IE program encountered several challenges, many of which were common across IEs. These generated a number of lessons learned to inform future IEs. First, the objective and intended use of the IEs was often not well defined. For example, it should be clear whether the goal of an IE is to help *determine a new USAID approach* to addressing a DRG program (i.e., a formative IE) or to *test USAID's existing approach* to addressing a DRG problem (i.e., a summative IE). The second set of challenges and lessons learned concern the decision to conduct an IE. Just because an IE was possible did not mean that one should have been carried out. For example, several IEs failed to test an intervention with an adequately robust theory of change. Furthermore, while academic PIs are generally regarded as a core strength of the DRG Center's IE approach, their role should match the goal of the IE, and it did not always make sense for the PIs to play a lead role in designing interventions. Third, inadequate IP buy-in, inadequate IP input, and conflict between evaluators and implementers accounted for most implementation challenges. Fourth, the DRG Center and its partners lacked strategies at the outset to move from a conflictive to a cooperative relationship between evaluators and IPs. Effective practices

included clear solicitation language, intensive post-contracting stakeholder engagement efforts, a weakening of the traditional firewall between evaluators and implementers, and assurance that the intervention is ready to be tested prior to the initiation of the IE. Fifth, the DRG Center and its partners also lacked strategies to ensure coordination and communication across IE stakeholders during implementation, including agreement on communication and information sharing protocols, active DRG Center and Mission engagement and facilitation, an in-country presence for evaluation teams, and an active role for DRG Center LPs (evaluation contractors).

IMPACT EVALUATION USE

There are several important examples of how DRG Center IEs have been used. The most salient of these was in Haiti, where the IE helped justify legal reforms and government funding for the legal defense of pretrial detainees. While we do find evidence of IEs informing existing projects, future projects, strategies, and general knowledge, there is considerable variation in IE usefulness. Several factors help explain this variation. Survey and case study evidence show that IE reports are often produced too late to inform decision making -- sometimes due to delays on the part of evaluators or the USAID Missions -- and at times due to idiosyncrasies related to the timing of other programs. Additionally, although there are good examples of dissemination, the survey suggests that reports were not widely distributed nor read on the whole. Furthermore, while evaluators were generally under the impression that reports were easy to read and contained actionable recommendations, IPs and USAID survey respondents were far less likely to agree. Finally, although post-evaluation action plans have been a USAID requirement since 2016, they were the exception rather than the norm.

WHAT SHOULD THE DRG CENTER'S APPROACH TO IMPACT EVALUATIONS BE GOING FORWARD?

This retrospective offers several key recommendations. On a broad level, it recommends that the DRG Center build from its previous IE program, rather than abandon the program or shift to an entirely different model. Nonetheless, the Center needs to implement some key changes. Among them, Missions and the DRG Center should make greater use of formal evaluability assessments, with an emphasis on defining the objective of a resulting IE, whether it be *formative* or *summative*, with more specific associated goals. Contracting should include a better-defined evaluation objective that clarifies stakeholder roles with specific provisions for IPs, evaluators, and academic PIs. In most cases, the conventional evaluation-implementation firewall should be dropped and instead stakeholders should work as an evaluation team, with a representative from the IP as an official team member, and a representative from the evaluators/PIs in-country for the life of the evaluation. In this process, Missions and the DRG Center could play a stronger role to ensure coordination and harmonization. Instituting these recommendations should encourage a much more nimble but far-reaching IE approach, and one that keeps a learning agenda at the fore. Emphasizing a clear IE objective, carried out by a well- coordinated evaluation team, would make possible more targeted dissemination and use both during and after a project. Dissemination and use would be further enhanced through increasing the accessibility and actionability of the findings report, involving USAID staff in crafting recommendations for Agency strategy and programming, and creating a central repository for posting research products.

1. INTRODUCTION AND RETROSPECTIVE QUESTIONS

In response to an influential 2008 National Academy of Sciences report, USAID's Democracy, Human Rights, and Governance (DRG) Center initiated a pilot program of IEs.[2] Initial IEs were done on an ad hoc basis through select IPs and existing mechanisms, in what we consider to be the first generation of DRG Center IEs. Based on the lessons learned from this experience, the DRG Center launched its own pilot mechanism to conduct IEs and other learning activities (Evaluating Democracy and Governance Effectiveness [EDGE; 2010–2014]), initiating a second generation of IEs. While evaluations in this second generation were still ongoing, the DRG Center scaled up and formalized its approach with the much larger DRG-Learning, Evaluation, and Research (LER) mechanism (2013–2022) and DRG-LER II (2018–present). In addition to a wide range of PEs, assessments, evidence and literature reviews, and other learning activities, the DRG Center and its LPs —particularly the NORC at the University of Chicago and Social Impact—have completed or are close to completing 27 IEs.

The DRG Center's IE initiative garnered substantial support among a core group of internal and external stakeholders, and the DRG Center has been at the forefront of USAID's overall efforts to assess the impact of its programming. Indeed, a recent study identified only 133 total USAID IEs published between 2012 and 2019, and only 72 of these met the formal USAID definition of an IE.[3] The DRG Center's 27 IEs, therefore, represent a significant portion of the Agency's total IEs.

Nonetheless, the DRG Center's IEs have also produced several critics frustrated with the challenging implementation process and concerned about IE usefulness. In 2019, the DRG Center began to scale back its IE work, and it initiated only two new potential IEs in 2019 and 2020. This retrospective intends to provide a look back at both the accomplishments and the challenges of DRG Center IEs with the goal of deriving lessons learned and providing evidence-based recommendations for future DRG Center evaluation work. This retrospective is also intended to serve as a lessons learned document for other donors, academic partners, and evaluators conducting IEs.

This retrospective answers the following five questions:

1. **Description**: How many IEs were initiated and how many were completed? What was the cost of these evaluations and what topics and regions did they target? What methodologies were used? For those that were not completed, why were they not completed?
2. **Findings**: At a high level, what has USAID learned from the findings of these IEs?
3. **Challenges and lessons learned:** What have been the challenges encountered in designing and carrying out IEs and what are the related lessons learned (for the DRG Center, Missions, implementing partners, and evaluators)?
4. **Use:** How has USAID (or others) used the IEs? Why were some evaluations more useful than others? How have findings been disseminated?

[2] Goldstone, Jack A., Larry Garber, John Gerring, Clark C. Gibson, Mitchell A. Seligson, Jeremy Weinstein (2008) *Improving Democracy Assistance: Building Knowledge through Evaluations and Research.* Washington DC: National Academies Press.

[3] Velez, Irene. (2020) *Assessing the Quality of Impact Evaluations at USAID*. Washington DC: USAID

5. **Recommendations**: What should be the DRG Center's approach to IEs moving forward? Under what conditions are they most effective and useful? How could the DRG Center better support Missions and others in the utilization of IE findings/recommendations?

LITERATURE REVIEW AND FRAMING

USAID's Automated Directives System (ADS) distinguishes between two broad types of evaluations. The first is IEs, which measure changes in development outcomes attributable to an intervention through the estimation of a credible and rigorously defined counterfactual.[4] IEs include both experimental evaluations, entailing random assignment of an intervention to beneficiaries, and quasi-experimental evaluations, in which a comparison group is purposively constructed. The second are observational evaluations, referred to as PEs), which include developmental evaluations, formative evaluations, some outcome evaluations, and process evaluations.

There continues to be a debate on the value of IEs.[5] There remain purists committed to randomized controlled trials as the only unbiased source of evidence.[6] Others are committed to IEs but believe that natural- and quasi-experimental methods can produce credible evidence about impact.[7] There are also strong IE opponents, including those who feel that the value of IEs is overstated or misleading.[8] To be sure, IEs have their limitations. As with all research, IEs are subject to sampling and measurement challenges. In addition, they also confront their own specific challenges, including errors in the randomization or matching process, non-compliance (i.e., inconsistency between treatment assigned and received), risks of spillover effects or other forms of contamination (i.e., control units receive treatment, or vice versa), and limits to external validity (i.e., the ability to generalize findings to other contexts). Moreover, there are many interventions that cannot be tested through an IE. For example, it might not be possible to identify a control group; the costs and challenges of an IE might outweigh the benefits; or a new intervention approach might be insufficiently consolidated and require extensive adaptation. Despite this debate, the majority of the literature recognizes that evaluations employing a rigorously defined counterfactual, whether that be through a randomized experiment or some other method, offer the best possibility for confidently estimating the impacts of an intervention.[9] The balance of scholarship agrees with USAID's ADS, which states, "When USAID needs information on whether an intervention is achieving a specific outcome, the Agency prefers the use of impact evaluations."[10]

---

[4] ADS 201.3.6.4

[5] See for example, Dawn Langan Teele (2014) Field Experiments and their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences. New Haven: Yale University Press.

[6] See for example, Alan S. Gerber, Donald P. Green, and Edward H. Kaplan. 2014. The Illusion of Learning from Observational Research. In Dawn Langan Teele ed. *Field Experiments and their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences.* New Haven: Yale University Press.

[7] See, for example, Dunning, Thad (2012) Natural Experiments in the Social Sciences: A Design-Based Approach. Cambridge: Cambridge University Press.

[8] See for example, Florent Bédécarrats, Isabelle Guérin, Francois Roubaud (2019), All That Glitters Is Not Gold. The Political Economy of Randomized Evaluations in Development. *Development and Change*, Vol. 50(3): pp.735-762.

[9] See for example: Esther Duflo and Michael Kremer (2005) Use of randomization in the evaluation of development effectiveness. In Pitman G.K, Feinstein O. N., & G.K. Ingram eds. Evaluating Development Effectiveness. World Bank Series on Evaluation and Development. Vol. 7. New Brunswick: Transaction Publishers, 205-231.

[10] 201.3.1.2

There have been several efforts by different donor and academic organizations to review the pitfalls and challenges of carrying out IE work. Within USAID, a review of the Bureau for Economic Growth, Education, and Environment's (E3) IEs identifies a number of challenges.[11] These include a lack of Mission buy-in, changes in LP contractors, inadequately defined interventions, poorly specified outcomes, failure to build on previous scholarship, inadequate adaptation to local context, and timing issues (IPs selecting sites or starting to work prior to randomization and baseline).[12] A review of "impact-oriented accompanying research" for the German Development Institute recommended carefully selecting topics for IEs, engaging researchers early on, clarifying expectations among stakeholders (e.g., researchers, practitioners, and IPs), determining the design collaboratively among stakeholders, communicating continuously, and viewing the IE as an opportunity for learning throughout the project cycle.[13] A broader review of IEs in German development cooperation called for increasing IEs, increasing financial resources, building capacity, creating incentives, involving the research community, building IE capacity in partner countries, and aggregating and using existing evidence.[14]

Beyond implementation challenges, both the World Bank and the Inter-American Development Bank report challenges in using IE findings.[15] Studies have found challenges in producing timely findings, generating actionable conclusions and recommendations, engaging decision-makers, and disseminating results.[16] A forthcoming study commissioned by the German Institute for Development Evaluation also finds substantial limits to dissemination and utilization, including a perception among practitioners that IEs are audits rather than learning exercises.[17] This report will show that the DRG Center has avoided some of these pitfalls. For example, it has done a good job of incentivizing IEs and providing funding for them; however, the DRG Center experience echoes many of these challenges.

METHODOLOGY

This retrospective uses a mixed-methods design to answer the retrospective questions. The design entails a broad comparison of all initiated and completed IEs and a deeper dive into eight case study IEs. The retrospective team used several sources of data. To look across all IEs, the team conducted a desk review,

[11] In 2020 this bureau was incorporated into a new Bureau of Democracy, Development, and Innovation along with the DRG Center.

[12] Molly Hageboeck, Jacob Patterson-Stein, Irene Velez (2017). *Opportunities for Enhancing Returns on E3 Bureau Investments in Impact Evaluations*. Washington DC: USAID; Molly Hageboeck, Jacob Patterson-Stein, Irene Velez (2019). Impact Evaluation: Critical Challenges/Promising Solutions. Washington DC: USAID. There is considerable overlap in the findings and recommendations of this study and these cited studies. Ours has a more robust methodology but arrives at many of the same conclusions. The main differences are that the DRG Center's IEs rely heavily on academic PIs, which is a central focus of this report, and the E3 bureau's experience was more impacted by a change in evaluation contractors.

[13] Evelyn Funk, Lisa Gross, Julia Leininger, Armin von Schiller (2018) Lessons Learnt from Impact-Oriented Accompanying Research: Potentials and Limitations to Rigorously Assessing the Impact of Governance Programmes. Bonn: German Development Institute

[14] German Institute for Development Evaluation (2019) Rigorous Impact Evaluation in German Development Cooperation. DEval Policy Brief 5/2019

[15] Independent Evaluation Group (2013) World Bank Group Impact Evaluations: Relevance and Effectiveness. Washington DC: The World Bank Group. Office of Evaluation and Oversight (2017) IDB's Impact Evaluations: Production, Use, and Influence. Washington DC: Inter-American Development Bank.

[16] Ibid.

[17] Aarti Mohan, Tobias Straube, and Surya Banda (Forthcoming 2021) Analysis of the Systematic Implementation of Rigorous Impact Evaluations and Evidence-Use in International Development Cooperation Organisations. Sattva Consulting and Scio Network.

an online survey offered to stakeholders in all previous DRG Center IEs, and KIIs. For the eight case studies, the retrospective team attempted to interview representatives of diverse IE stakeholders, each of which are discussed in turn below. Because data have not been tracked systematically for the DRG IEs, beyond what could be reconstructed from document review, the team necessarily draws heavily on the perception-based interviews and surveys. The evaluation team analyzed quantitative survey and qualitative interview data, cross-referencing findings against each other to enhance validity and to mitigate limitations of different data sources.

Desk Review: As part of the background research, the team conducted a desk review of relevant policy and background documents. This included a review of best practices from other organizations that conduct a high volume of IEs, including the World Bank, Millennium Challenge Corporation (MCC), Innovations for Poverty Action (IPA), and the Abdul Latif Jameel Poverty Action Lab (J-PAL). The team also conducted a document review and coding of all completed DRG IE reports. The coding captured basic meta-data on the evaluation, details on the evaluation methodology, a coding of findings, and mentioned methodological and implementation challenges among other factors.

Case Study Selection: To focus the retrospective, the team identified eight completed IEs that represent key dimensions across the range of IEs that were implemented during EDGE, DRG-LER I, and LER II. Case studies were selected purposively to ensure variation in 1) the robustness of the theory of change, 2) challenges in the implementation process, 3) results (i.e., positive and null/negative), and 4) utilization (i.e., use or lack thereof). With four different variables of interest and many other sources of variation, it was not possible to select cases in such a way that would allow for meaningful control.

CASE STUDY IES

Countering Violent Extremism in Bangladesh

Constituency Dialogues and Citizen Engagement in Cambodia

Evaluation of Secondary Prevention in the Community, Family, and Youth Resilience Program in St. Lucia, St. Kitts, and Nevis and Guyana

Ghana Strengthening Accountability Mechanisms (GSAM) IE

Governance, Accountability, Participation, and Performance (GAPP): short message service (SMS) for Better Service Provision in Uganda

Impact Evaluation of USAID Haiti PROJUSTICE Program Pretrial Detention Component

Impact Evaluation of USAID/Malawi Local Governance Accountability and Performance (LGAP) Activity

Media and Civil Society in Tanzania

Interviews and Group Discussions: The authors conducted interviews and group discussions from January 4–February 22, 2021. Key informants include personnel from other institutions conducting IEs (e.g., World Bank, Department of Labor); current and former DRG Center staff; and staff from the Bureau of Planning, Policy, and Learning. In addition, for each of the case study IEs, the research team sought the perspectives of principal investigators (PIs), DRG Center LPs, relevant USAID Mission staff, and IP staff.

Accounting for both individual interviews and group discussions, our qualitative data include perspectives shared by 64 individuals. Please refer to Table 5 in Annex 2, KII Interviews, for a summary table of the number of interviews and interviewees by IE and stakeholder type.

Online Survey: The research team conducted a quantitative survey with individuals across all key stakeholder groups for each DRG Center IE. Individuals who were adequately involved in the IE to speak knowledgeably about the process, findings, and use were included in the survey target population. LPs provided initial names of stakeholders for each evaluation and these individuals were asked to identify additional stakeholders. This approximately 30-minute online survey was distributed via Qualtrics between December 22 and February 26 and included questions about IE use, lessons learned, and recommendations. It employed a conjoint experiment, in which the evaluation team presented combinations of IE characteristics for two hypothetical IEs and gauged which specific components of IEs respondents preferred. The team employed this method because there is no variation within the 27 DRG IEs for some characteristics (e.g., nearly all DRG IEs were randomized) and so the team could carry this out within hypothetical profiles of IEs. Of 127 individuals invited to respond to the survey, 80 participated, yielding a response rate of 63 percent. At least one response was received for 22 of the 27 IEs for which respondents were recruited. The evaluation team also invited individuals that were part of IEs that ultimately did not move forward, but had a very low response from this population. Please refer to Table 1 for a breakdown of quantitative survey respondents by stakeholder group.

TABLE 1. SAMPLE BY DATA COLLECTION METHOD AND STAKEHOLDER GROUP

STAKEHOLDER GROUP	KII	QUANTITATIVE SURVEY	CLASSIFICATION FOR SURVEY ANALYSIS
DRG Center staff or former staff	4	20	USAID
USAID Mission staff involved in IEs	12	3	USAID
Evaluator	14	11	Evaluator
PI	17	19	Evaluator
IP	10	15	Implementer
Other IE practitioners	7	0	NA
Total	**64**	**80**	

Data Analysis: The evaluation team used multiple techniques to analyze the data. For the survey data, the team examined descriptive statistics (e.g., means, crosstabs) to obtain statistical profiles of the samples and to explore differences by stakeholder group. The team analyzed the results of a conjoint survey experiment consistent with a pre-analysis plan registered with the University of Texas at Austin. For the qualitative data analysis of interviews, group discussions, and qualitative survey responses, the team used simple thematic coding. The team identified broad themes, both deductively based on the retrospective questions and inductively based on interviews, and organized qualitative data by these themes in a spreadsheet. Findings were derived by comparing across data sources within each theme category.

For both qualitative and quantitative data, the team disaggregated findings along relevant dimensions, such as stakeholder group and IE, to conduct subgroup analyses and identify trends by stakeholder group.

Risks and Limitations: There are a few methodological limitations worth noting. First, it was not possible within the constraints of the retrospective to conduct detailed qualitative data collection on all 27 IEs, and while the eight case studies allow for depth, there are likely some lessons that were missed in the remaining 19 IEs. The review of IE reports and the survey, with responses for 22 IEs, helped mitigate this risk. Second, while the interviewing produced a wealth of qualitative data, given the subjective nature of perception-based data and the contentious nature of some IEs, different stakeholder groups on the different IEs often disagreed with one another and it was not always possible to reconcile these accounts to identify basic facts about an IE experience. Third, in both the survey and the interviews, there was somewhat over-representation of evaluators and PIs vis-a-vis USAID Mission staff and implementers. The team attempted to mitigate this concern with the quantitative data by analyzing by subgroup rather than pooling the data, and with the qualitative data the team was careful to give extra consideration to IP perspectives.

2. DESCRIPTION OF THE DRG CENTER IE PROGRAM

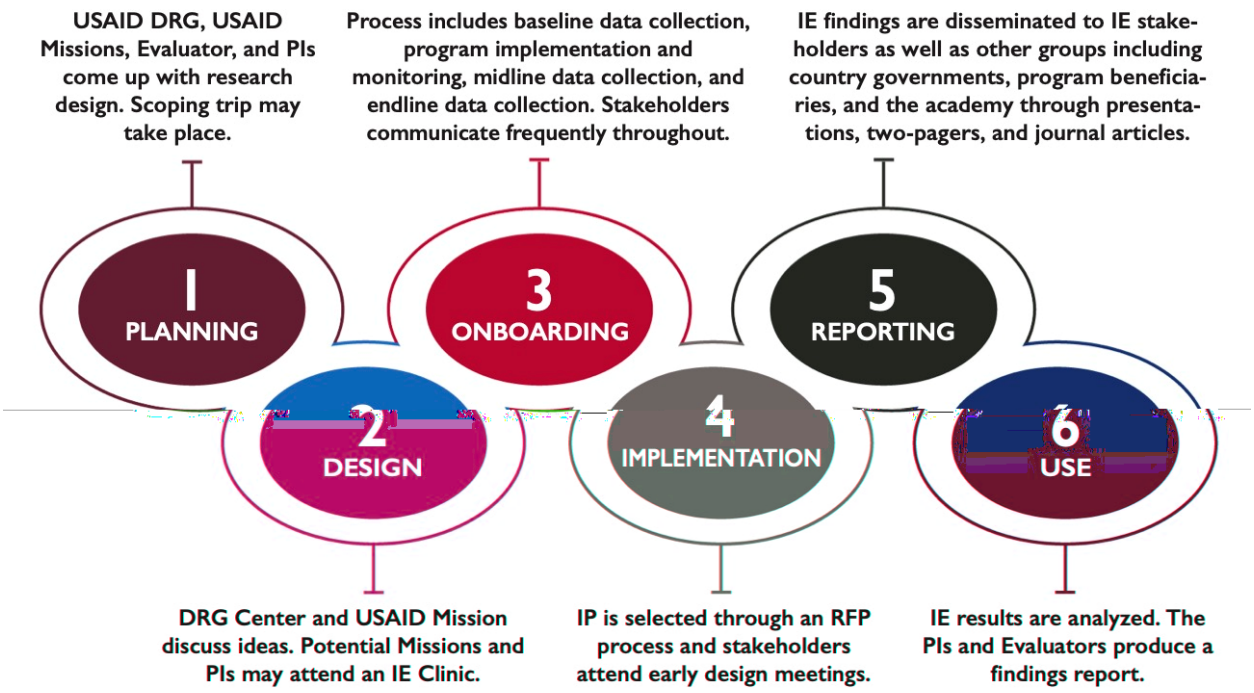
Question 1: Description: How many IEs were initiated and how many were completed? What was the cost of these evaluations and what topics and regions did they target? What methodologies were used? For those that were not completed, why were they not completed?

The DRG Center Model: The DRG Center's approach to IEs evolved over time. The initial generation of IEs was done on a somewhat ad hoc basis by encouraging Missions and IPs to develop IEs of their programs. Interviews suggest that USAID derived several conclusions and lessons learned from these initial efforts: 1) Missions lacked the expertise, the incentive, and the buy-in to conduct IEs on their own; as such, the DRG Center needed to play a proactive role in promoting, incentivizing, and supporting IEs. 2) Academic partners offered a means to supplement the expertise that USAID, IPs, and even LPs lacked. The Evidence in Governance and Politics network specifically offered a well-qualified network of potential academic PIs. 3) IPs were willing to conduct internal IEs; however, they lacked adequate capacity and, more importantly, they did not have incentives to publicize null or negative findings. As such, IEs needed to be conducted by external evaluators. 4) External IEs needed to be planned at the activity design stage and prior to procurement and award to an IP.

The DRG Center learned from these early experiences and developed a unique and ambitious approach that included several core elements: 1) The DRG Center created the highly flexible EDGE, and later DRG-LER I and LER II mechanisms, to commission IE work, and in 2012 it created the Learning Division to operate and oversee these mechanisms. 2) Along with its LPs NORC and Social Impact, the DRG Center cultivated ties to academics not only to lead IEs but also to actually design interventions based on theory and evidence. Several PI interviewees praise DRG Center staff for their ability to understand and work with academics. 3) The DRG Center launched an impressive effort to build Mission buy-in and capacity through outreach, training, and co-funding support to Missions. Most importantly, between 2013 and 2017, the Learning Division ran IE clinics whereby academics, DRG Center staff, and Mission staff could come together to learn about IEs, discuss evidence, and design future IEs prior to the procurement of an

intervention.[18] While IEs took many different routes, Figure 1 shows the various stages of a typical DRG Center IE.

Figure 1: Illustration of the IE process.



SOURCE: AUTHORS BASED ON INTERVIEWS

Completion: To date, the DRG Center and its LPs have completed or are close to completing 27 IEs. The IE pipeline is something of a funnel. Any Mission can express interest in an IE, but many of these initial conversations did not move past an early screening by the DRG Center. Those that did pass moved to further consideration, which included but was not limited to Mission attendance at an IE clinic. From these, the DRG Center figures suggest that 42 evaluations moved on to the design stage. Of these, 33 continued on to baseline data collection, and 9 did not make it past the design stage. Two were cancelled after baseline and one was converted into a long-term PE, leaving 30. Of these, 25 are complete, two are very close to complete, two are in progress and will be completed in 2021 or 2022, and one is at the design phase. Among the completed or close to complete IEs, five were part of the initial generation of IEs conducted with some DRG Center support but prior to the initiation of the EDGE or DRG-LER mechanisms. Annex 1 contains a listing of the IEs analyzed in this report.

These IEs vary widely in terms of cost, subject matter, and geographic area covered.

IE Cost: Figure 2 shows the cost of each IE by region and sector. The figure excludes those IEs for which the team does not have full budget data, including the first generation of IEs and IEs where the DRG Center only supported data collection. The IEs funded by the DRG Center have a wide range of costs,

[18] The IE clinics were highlighted as a best practice in Hageboeck et al. Impact Evaluation: Critical Challenges/Promising Solutions.

from $230,000 to just over $2.5 million with a mean cost of $713,202 and a median cost of $557,582. Factors that drive cost include the length of the evaluation, scope of research questions, and country-specific data collection costs. The $2.5 million evaluation occurred in three Caribbean countries, each with high data collection costs. Several African and Latin American countries also had high data collection costs, while Asian countries tended to have lower data collection costs.

Figure 2. IE Costs by Region

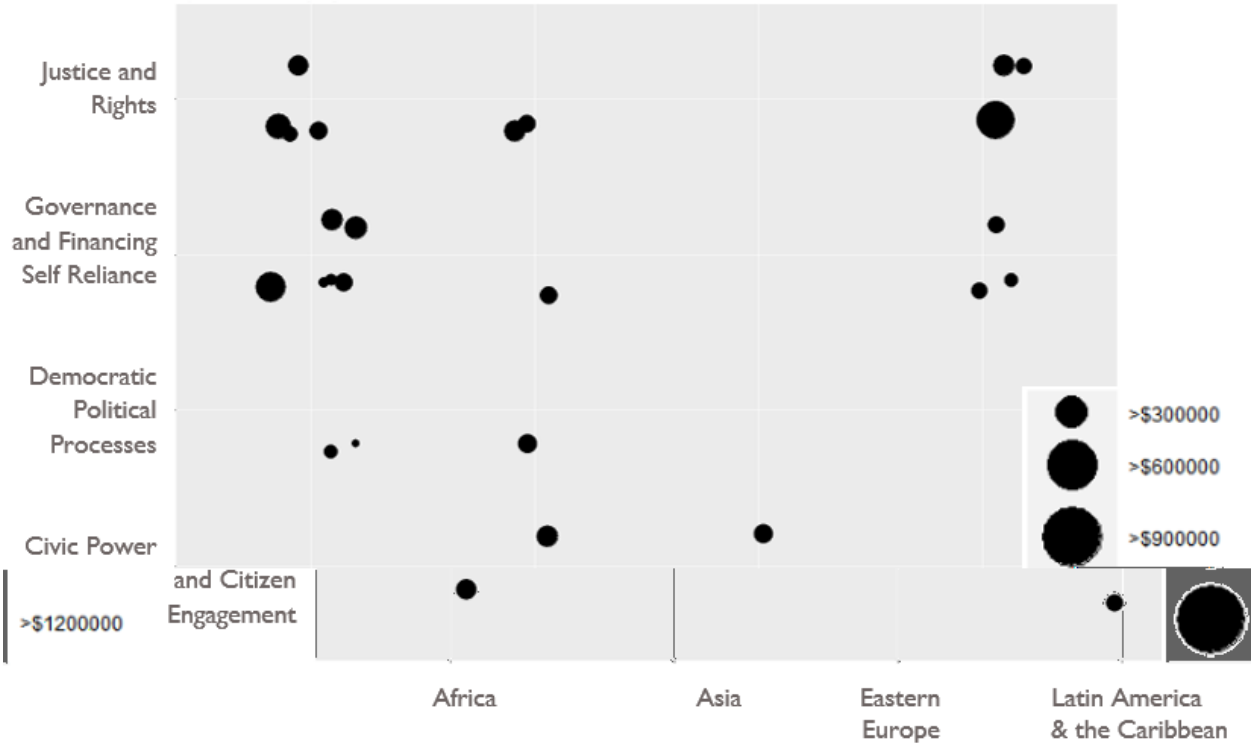


Table 2 presents data from USAID's evaluation registry and compares DRG and all USAID evaluations completed between 2016 and 2019. As shown in the table the DRG Center's mean and median costs are fairly consistent with USAID's experience overall. It is difficult to compare USAID IE costs with other institutions. A 2017 Inter-American Development Bank report attempted a cost comparison, but its estimates of its own IEs appear to exclude the time of staff carrying out the evaluation.[19] The report does suggest that USAID evaluations are cheaper than MCC, on par with the World Bank, and more expensive than J-PAL.[20] As shown in Table 2, the average IE tends to be a little more than three times the cost of the average PE, and the median IE tends to be a little less than three times the cost of the median PE. The costs are in some ways surprisingly low—given the extended timeframe involving evidence reviews, (variable) scoping, design, baseline, midline (variable), and endline—compared to the one-shot nature of a typical PE that involves one 3–4 week period of qualitative field work. The price paid by USAID for the IEs was actually considerably less than the true cost, however, as interviewed academic PIs reported substantial underbilling of their time. In addition, in the case of several IEs (e.g., Mozambique, Ghana,

[19] Office of Evaluation and Oversight. IDB's Impact Evaluations: *Production, Use, and Influence*.

[20] *Idib.*

Uganda) academic PIs paid for some IE activities out of their university research budgets or complementary grants.

TABLE 2: MEAN AND MEDIAN COSTS OF REGISTERED DRG AND USAID PES AND IES (2016–2019)

	DRG		ALL USAID	
	Performance Evaluations	**Impact Evaluations**	**Performance Evaluations**	**Impact Evaluations**
No. of evaluations	146	11	624	65
Average evaluation budget	$ 212,946	$ 701,876	$ 265,999	$ 989,336
Median evaluation budget	$ 170,000	$ 500,000	$ 200,000	$ 558,000

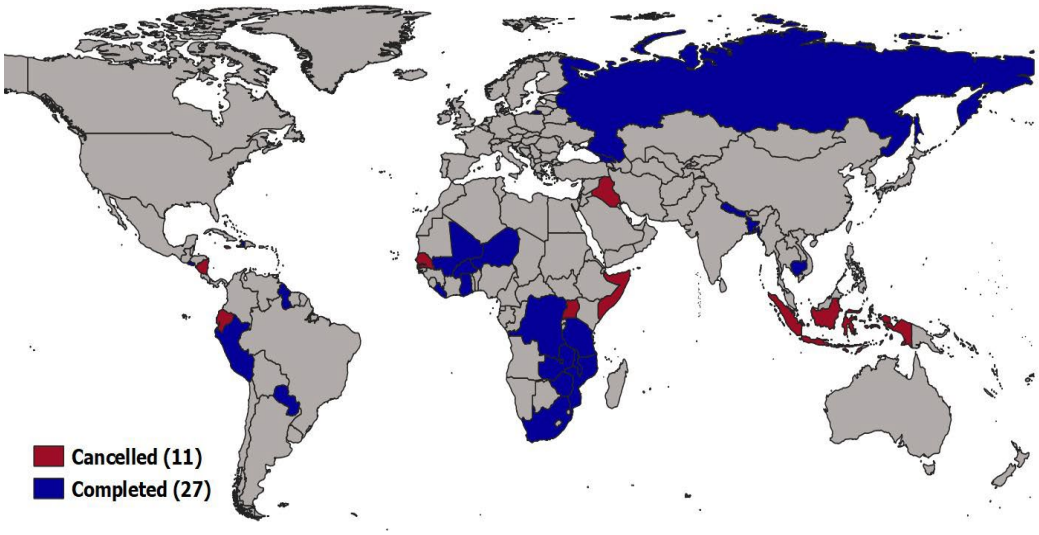
SOURCE: USAID Evaluation Registry Dashboard

NOTE: While all evaluations are required to be included in the registry it is possible that some are excluded.

IE funding under EDGE, DRG-LER, and LER II comes from two sources: USAID Missions and the DRG Center. In most cases, the Learning Division would contribute half of the total costs up to $250,000 with remaining funding provided by the USAID Mission. As noted below in response to Question 3, DRG Center co-funding was a critically important element of the DRG Center's approach that helped defray the costs to Missions and bring the cost of IEs closer to the cost of PEs. It is difficult to assess the value for money of IEs; however, if a Mission is able to take advantage of the Evidence Review in its design and utilize baseline data to inform implementation, and USAID and others are able to use the final results, then the return on the investment would be very much worth it. As discussed in the response to Question 4, this was not always the case.

Region: The region with the most IEs was Africa, with 15 in progress or complete. Within Africa, most IEs were in Southeastern and West Africa, with Missions in the Democratic Republic of the Congo and Uganda each conducting multiple IEs. Additional IEs were conducted in Burkina Faso, Ghana, Liberia, Malawi, Mali, Mozambique, Senegal, Somalia, South Africa, Tanzania, Zambia, and Zimbabwe. Latin America had the next largest number of IEs, at seven. These were dispersed throughout the Caribbean (Haiti, the Dominican Republic, St. Lucia, St. Kitts and Nevis, and Guyana), Central America (Nicaragua), and South America (Peru and Paraguay). Asia conducted six IEs concentrated in Cambodia (4), Nepal (1), Indonesia (1), and the Middle East (Iraq and the West Bank). The two IEs in the Eastern Europe/Eurasia region were conducted in Georgia and Russia.

Figure 3. DRG IEs by Country



Cancelled IEs: Not every program or idea is a good fit for an IE, and program leadership varies in its willingness to pay the associated costs. As noted above, the DRG Center filtered out proposed IEs that were not a good fit. Those that were a good fit were often invited to participate in an IE clinic to help refine the idea. Of the nine IEs that were cancelled in the design stage, the reasons for cancellation varied. In Ecuador and Nicaragua, IEs were cancelled with USAID Mission departure and presence reduction. In Senegal, the cancellation was attributed to a change in Mission leadership. In Jamaica, PIs and the Mission were unable agree on an intervention to test. In one evaluation in Uganda, the PIs recommended not going forward with the IE due to an inadequate theory of change.

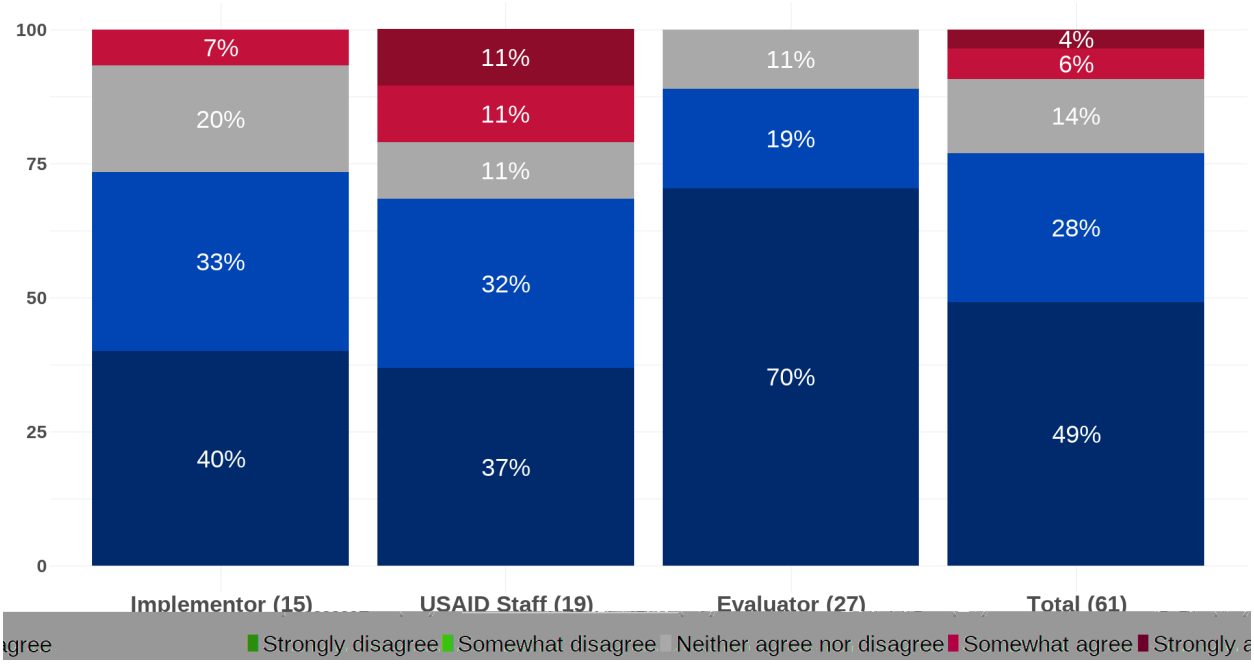
Only two IEs were cancelled after baseline data collection. The first was an evaluation of a peacebuilding project in Guatemala, which was canceled after baseline data collection. Reasons cited in interviews included budget cuts to the Guatemala Mission, Mission staff turnover, and IP concerns about aspects of program implementation. The second was an evaluation in the West Bank and Gaza that was canceled at midline when the USAID Mission ceased activities. A third IE, on rule of law in the Caribbean, was changed from an IE to a rigorous PE due to data collection challenges and budget cuts to the Mission; all stakeholders were in agreement with the change.

The DRG Center experience demonstrates that even with a robust screening process, not all planned IEs will go forward; in this case, a little less than one-third did. Several key informants felt that this was a good thing—given all the challenges of conducting an IE, it can be preferable to halt an IE rather than push forward a bad fit. Others raised concerns that a Mission's ability to reverse course at any point during the IE process as their priorities shift presented a vulnerability to the success of the DRG Center's IE portfolio.

Achievements: The DRG Center's IE program was a model in creating academic-Mission linkages to implement a robust IE program. Through outreach, training, co-funding, and the clinics, the DRG Center created Mission demand to conduct IEs, obtained buy-in, and clearly demonstrated that IEs in DRG were possible. The IE program attracted top academic talent to serve as PIs, who were able to produce and implement strong IE designs. Furthermore, the DRG Center's IE program avoided some of the pitfalls that befell other USAID IE initiatives. For example, it appeared to achieve a higher level of Mission buy-in and

enthusiasm and avoided contracting problems of changing evaluation contractors mid-way through evaluations.[21] Surveyed stakeholders seemed to recognize the value of IEs. Of respondents who answered the survey question, 78 percent agreed or strongly agreed that they would encourage someone else from their organization to participate in an IE, including majorities in USAID and among IPs.

Figure 4. Respondents Would Encourage Others at Their Organization to Participate in an IE



In summary, the DRG Center's approach incorporated several lessons learned from an initial generation of IEs to build a successful IE program that entailed a flexible contracting mechanism, the involvement of top academics, and mechanisms to build Mission buy-in, including training, IE workshops (clinics) and co-funding. Through its efforts, the DRG Center has supported 27 completed or close-to-complete IEs with three more in process. These IEs covered a range of DRG issues and were geographically dispersed. Several IEs did not move past the design stage for a variety of reasons; only two IEs had to be cancelled after baseline data collection. Of those for which the team has data, the median IE cost $557,582 and the average cost $713,202, which is on par with other USAID offices and other IE contracting organizations. In many ways, the DRG Center's IE program was a model in creating academy-Mission linkages to implement a robust IE program.

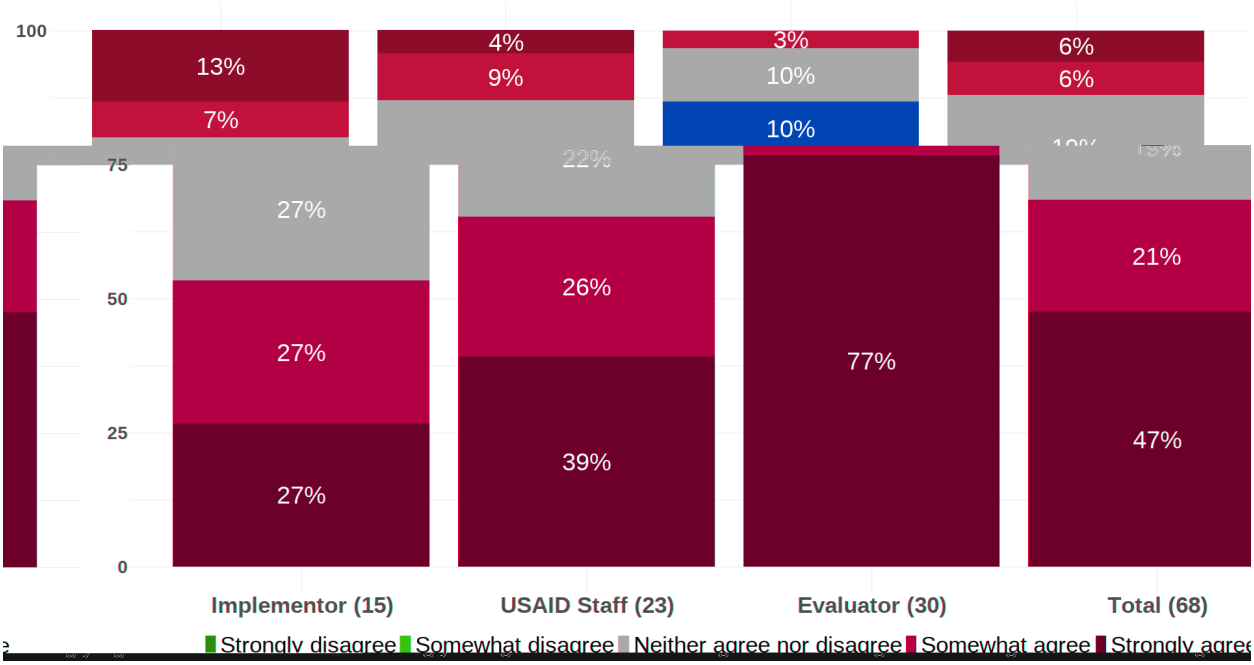
[21] Molly Hageboeck, Jacob Patterson-Stein, Irene Velez (2017). *Opportunities for Enhancing Returns on E3 Bureau Investments in Impact Evaluations*. Washington DC: USAID.

3. FINDINGS OF IMPACT EVALUATIONS

Question 2: What has USAID learned from the findings of these IEs that would not have been learned through a PE, monitoring and evaluation (M&E), or other methodology?

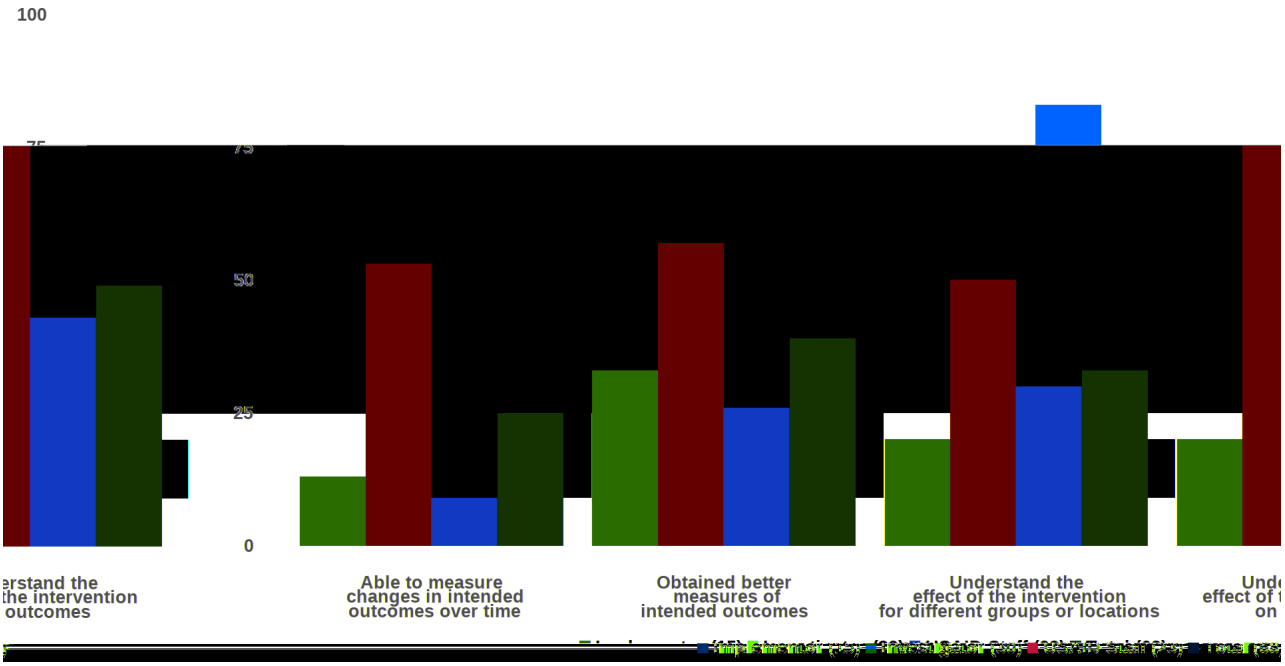
To assess the value of IE findings, the evaluation team asked survey respondents to compare IEs with traditional monitoring and evaluation. Of all survey respondents, 47.5 percent strongly agreed and another 21 percent agreed with the following statement: "We learned more from the project IE than could have been learned from more typical monitoring and a performance evaluation." Only 12 percent either somewhat or strongly disagreed. This does vary significantly by stakeholder group, however, with roughly 90 percent of evaluators in agreement, but only just over half of IPs in agreement. This difference in perspective is likely linked to both understanding of the methodology and how each stakeholder values IEs, with evaluators more convinced of their value than implementers.

Figure 5. Percentage of Respondents who Learned More From an IE Than From Other Monitoring, Evaluation, and Learning (MEL) Tools



More specifically, the survey asked respondents what they learned from the IE that they could not have learned from a PE or from project monitoring. As shown in Figure 6, the most common advantages of IEs cited by respondents were that the IE helped understand the project's effect on outcomes, as well as the effect of the intervention for different groups of locations. Evaluators and PIs were more likely to report that the IE contributed to unique learning across all four categories, and USAID stakeholders were least likely to say the IEs obtained better understanding of intended outcomes or were able to measure changes in intended outcomes over time. However, in KIIs, USAID personnel frequently mentioned that IEs provided evidence that program dollars were being well spent and that programs were effective in a way that MEL tools alone could not. Responses to the open-ended questions mentioned the value of a counterfactual, as well as an IE's ability to understand the mechanism behind the program theory of change.

Figure 6. Learning From IEs, by Stakeholder



A few examples of potentially impactful IE findings are highlighted in the sidebar on the following page. For more information about the findings of specific IEs, please refer to Annex 1: Learning Harvest, and note that both positive and null findings are potentially valuable to USAID. In the case of Haiti, an IE found that pretrial detainees in Haitian prisons who were provided with legal support had their cases brought to conclusion prior to those that did not receive legal aid. Moreover, the IE included a cost analysis that found that legal support was cheaper to the Haitian state than detaining the accused. These findings could not have been determined with the same confidence using other evaluation approaches. In the case of the Caribbean, USAID funded an implementer to export a promising family counseling-based approach targeting youth at risk for involvement in crime and violence. Longitudinal studies of its effectiveness conducted by the implementer suggested that it was enormously successful in reducing risk factors among at-risk youth. However, when tested with an IE, the evaluation found that the tool used to measure change in risk over time produced almost identical results in a control group as it did in the treatment group.

Although these two evaluations produced important findings that could not have been learned through PEs or activity monitoring, in many cases evaluation findings were less conclusive or yielded less learning for practitioners. As one USAID interviewee noted, “All that time and effort, and I can find very little of value in the final report." Concerns included the following:

- In Tanzania and Mali, for example, researchers were not sure whether the null findings meant that the intervention was ineffective as it was designed, if it was implemented poorly, or if there was some other factor at play.
- Many studies produced mixed results, whereby some tested variables changed due to the intervention but others did not. These evaluations did not produce the clear policy guidance for which some practitioners might have hoped. IEs in Ghana and Malawi are strong examples of this.
- In some cases, particularly when interventions were rooted in a weak theory of change, one likely did not need a control group to conclude that at least one of the treatment arms was not going to produce the desired outcomes (South Africa, Peru, Zimbabwe). For example, one arm of a multi-armed Randomized Control Trial (RCT) that includes only a training intervention could be expected not to impact the key outcomes of interest, but may have been included in the design.
- Some implementers felt that their interventions were too hamstrung by an evaluation design that needed to randomize, constrain the timing of activities, limit spillover effects, and otherwise overly standardize to really test the intervention as they felt it should be implemented (South Africa, Caribbean, Georgia, Paraguay).
- Often, there were limits to what an evaluation could test. For example, the IEs were typically adequately powered to test overall impact, but they typically did not have enough observations to test whether their effect varied among different groups.

As such, not all IEs produced clear, actionable findings about impact. Even in these cases, however, IEs still generated potentially valuable data. For example, in the Cambodia-National Democratic Institution (NDI) evaluation, the baseline provided the implementers with data on voter knowledge and participation that was previously unknown to them. In Uganda, when midline data suggested that the intervention, an SMS platform that allowed residents to lodge complaints about public services to district officials, was not going to have an impact, the evaluation team shifted gears and re-programmed funds to explain variation in use of the platform. The team's analysis provided compelling evidence about village-and individual-level factors that accounted for uptake of the platform, and the IP incorporated the evaluation recommendations into their activity design.

In summary, unlike more traditional PEs and monitoring, IEs

SNAPSHOT OF IE FINDINGS

In Zambia, the IE found that citizens use information on candidate performance to vote—rewarding high performers and punishing low performers. The citizens' decision to vote was not affected by the ethnic background of the candidate when performance information was available. Information on candidate performance compared to benchmarks was useful to citizens for activities unfamiliar to them.

A policy dialogue activity in Nepal found generally positive but modest and mostly short-term impacts of screenings of candidate debates and small-group discussions hosted by community radio stations ahead of federal parliamentary elections. These activities improved participants' sense of political efficacy and swayed participants' views about the role of government. The activities also influenced what issues were discussed by the candidates taking part in the debates.

An accountability program in Ghana that employed both a top-down and bottom-up approach was found to have some important impacts, even though many features of district governance did not change. The program's bottom-up civil society organization (CSO) programming had more effect on citizen engagement and the behavior of administrators than top-down GAS programming. On the other hand, the top-down programming had a larger effect on politicians, who decreased political manipulation of the budget.

In Peru, an anti-corruption program had no effect on the awareness of local political corruption, personal experiences with corruption, or attitudes toward corruption. However, survey experiments showed that when respondents are confronted with corruption and bad performance separately, they tend to punish corrupt politicians and bad-performing politicians at similar rates. The IE also found that when it comes to bribery, Peruvians are especially willing to report offending public officials when they do not deliver on their promises to provide administrative favors or benefits of some kind.

were able to measure a counterfactual for an intervention and make causal inferences about that activity's impact. In Haiti, an IE demonstrated that an activity was working and should be scaled up. In the Caribbean, an IE found that a previous M&E's conclusion that an intervention was producing dramatic results was incorrect. Furthermore, IEs frequently provided better measures of outcomes and changes in those outcomes over time than traditional M&E. This is likely because PIs and evaluation teams have greater expertise in developing outcome indicators for hard-to-measure concepts like accountability than most MEL Activity staff, and can devote more resources to developing data collection systems than most MEL Activity budgets allow. The nature of IE outcome indicators also allows for greater freedom and creativity than MEL indicators, since they do not have the same requirements for documentation, and there are no targets to achieve.

IEs also offer opportunities for richer data collection and analysis through the life of the activity. Though not exclusive to experimental evaluations, baseline data or regression analysis conducted with IE data produced valuable information that could be useful for implementers and could provide additional opportunities for learning before the final IE results are available. However, many evaluations did not meet stakeholders' expectations. USAID and IP survey respondents were considerably less likely than evaluators and PIs to identify the benefits of IEs. In some cases, null results could not be explained, mixed results did not produce clear policy guidance, evaluations of weak interventions produced little value added, and there were limits to what an IE could test.

4. LESSONS LEARNED

Question 3: Challenges and Lessons Learned: What have been the challenges encountered in designing and carrying out impact evaluations and what are the related lessons learned (for the DRG Center, Missions, implementing partners, and evaluators)?

In this section, the evaluation team explores the challenges and lessons learned in conducting DRG Center IEs. These are grouped into five big picture lessons learned: (1) Clarifying the purpose of the IE is critical to all other evaluation activities, including the design, stakeholder engagement, conclusions, dissemination, and use; (2) Successfully initiating an IE requires many pieces to fall into place; (3) Inadequate IP buy-in, inadequate IP input, and conflict between evaluators and implementers accounted for most implementation challenges; (4) the DRG Center and its partners lacked strategies at the outset to move from a conflictive to a cooperative relationship between evaluators and IPs. Effective practices included clear solicitation language, intensive post-contracting stakeholder engagement efforts, a weakening of the traditional firewall between evaluators and implementers, and assurance that the intervention is ready to be tested prior to the initiation of the IE; (5), the DRG Center and its partners also lacked strategies to ensure coordination and communication across IE stakeholders during implementation, including agreement on communication and information sharing protocols, active DRG Center and Mission engagement and facilitation, an in-country presence for evaluation teams, and an active role for DRG Center LPs.

IEs ARE NOT CLEAR IN THEIR OBJECTIVE AND INTENDED USE

While evaluators tend to focus on the methods used in IEs, the objective and use of an IE can also vary substantially. Interviews with stakeholders often revealed uncertainty and disagreement about the goal and the use of a given IE. Throughout this section, the team argue that a lack of clarity on the objective

and intended use produced several negative consequences. The following paragraphs lay out a typology of IEs based on purpose and use.

Evaluations are typically divided into *formative* evaluations and *summative* evaluations. Although definitions vary across sources, there is general agreement that a formative evaluation is designed to aid in developing an approach and a summative evaluation is designed to test if that approach worked. While an IE is often considered a summative evaluation method, the originator of the terms notes that the intended distinction is between how an evaluation is to be used rather than how it is to be carried out.[22] As such, donors like USAID can use IEs to help determine its approach to addressing a DRG activity (formative IE), or they can use IEs to test their approach to addressing a DRG problem (summative IE). Some of the IEs studied were more oriented toward the former objective and some more oriented towards the latter. Within the broad category of formative IEs the team identified *Innovating IEs*, *Pilot-Scale IEs,* and *Complementary IEs*. Within the category of summative IEs the team identified *Confirmatory IEs*, *Generalizing IEs*, and *Optimizing IEs*. Within USAID's DRG space, not all of these types have been used; however, they likely would (or should) be used to the extent that IEs continue and increase.

FORMATIVE IES

Innovating IEs: A first approach is to identify promising solutions to DRG problems that USAID does not yet know how to address. This was the case in Bangladesh, where the Mission wanted to confront a somewhat new threat of violent extremism. PIs conducted a lengthy evidence review and proposed testing a bystander model, which was piloted and tested and showed promising results.

Pilot-Scale IEs: A second related approach is to evaluate a new activity, or pilot, to determine whether to take the pilot to scale. In fact, the ADS requires USAID operating units to conduct IEs of pilot initiatives (ADS 201.3.6.5), although it appears that this requirement is frequently avoided.

Complementary IEs: A third formative evaluation approach is to learn about impact as a secondary objective of programming's core activities, or as part of a monitoring or a PE. In cases in which no evaluation is commissioned, or a monitoring or a PE is conducted, the Mission and IPs should arguably always seek to learn about the effects of interventions. For components of projects, it is often possible to identify comparison units with little extra effort, such as through the use of government administrative data, and in ways that IPs, the Mission, and perhaps performance evaluators, could still learn important lessons about intervention effects to aid in better programmatic decisions more generally. Complementary IEs are outside the empirical scope of this retrospective but remain an important option in a broader typology of IEs and a broader learning mindset across all programming could help build the broader knowledge base that underlies programmatic decisions.

[22] Scriven, Michael. "Beyond formative and summative evaluation." Evaluation and education: At quarter century 10, no. Part II (1991): 19-64. By contrast, the GAO more recently defined a summative evaluation as one determining "the extent to which a program has achieved certain goals, outcomes, or impacts," essentially synonymous with an IE. Our distinction would be consistent with the GAO definition if the term "program" is seen as an overall approach taken by a donor rather than as an individual activity carried out by an IP. GAO (2021) Program Evaluation: Key Terms and Concepts. GAO-21-404SP. Government Accountability Office: pg. 5.

SUMMATIVE IEs

Confirmatory IEs: Turning to summative approaches, a fourth reason to conduct IEs is when USAID and its partners continue to implement, and even replicate, interventions without evaluating their impact. This was the case in Cambodia, where NDI had been conducting community dialogues for many years but there was uncertainty about their impact. It was also the case in the youth violence prevention sector, where a family counseling-based approach appeared promising, and was being exported to other contexts (i.e., the Caribbean), but had not been adequately tested. Confirmatory IEs have been helpful in the broader assessment of community-driven development and reconstruction programs, yielding useful feedback about their utility.[23]

Generalizing IEs: A fifth approach is to use an IE to evaluate the effectiveness of a project in a new context. In this case, an IE has already demonstrated the effectiveness of an intervention in one location or among one population, but there may be context-specific factors that led to its success. As such, a generalizing IE tests the intervention, oftentimes somewhat adapted to the context, in a new setting, among a new population, or on a somewhat different outcome. To some extent, this was the case in Ghana, where the IE tested citizen scorecard campaigns that had been shown to be effective elsewhere.

Optimizing IEs: A sixth approach is to use IEs to optimize an intervention. Optimizing an intervention could be based on a number of criteria and here the team notes one. An IE might show that an intervention is effective, but there remains the possibility that a similar, but less costly, intervention exists. *Optimizing IEs* alter an intervention along one core dimension, such as price, and then systematically test the variation in that dimension to optimize on investment and outcome. As with Complementary IEs, in the formative evaluation approach above, the evaluation team is not aware of DRG Center IEs that have conducted *Optimizing IEs*. But, moving forward (especially if USAID and the DRG Center continue to carry out IEs), these *Optimizing IEs* will be an important summative evaluation tool. To provide one example of an *Optimizing IE* in another context, interventions offering bed nets at several different consumer price points, including free distribution, helped donors understand which of the approaches to providing bed nets worked best, and contrary to predictions, providing them for free led to more and better use.

As will be discussed below, a lack of clarity on the goal and use of the evaluation created several challenges.

SUCCESSFULLY INITIATING AN IE REQUIRES MANY PIECES TO FALL INTO PLACE

As discussed above in the response to Question 1, the DRG Center learned several lessons from the first generation of IEs, and it played a key role in attracting Mission interest and working with evaluation and academic partners to move the IE program forward. The evaluation policy, training efforts, co-funding, Mission champions, a willing pool of potential academic PIs, the IE clinics, and the evaluation mechanisms discussed in Question 1 all served as key factors in the development of the DRG Center's IE program. Nonetheless, initiating the DRG Center IEs still encountered numerous challenges.

[23] Casey, K. (2018) Radical Decentralization: Does Community-Driven Development Work? Annual Review of Economics 10:1, 139-163

Mission demand outweighed a desire for a learning agenda: Following the first generation of IEs, the DRG Center aimed to conduct IE work on political participation and local government accountability.[24] These were selected as 1) they were commonly implemented interventions, 2) there was uncertainty about what worked, and 3) they offered the possibility of randomization. Although several IEs have fallen within these categories, the DRG Center quickly branched out beyond this learning agenda in response to Mission demand and IE champions in the Missions. In fact, in most cases, the motivation to conduct an IE was not to fill a priority evidentiary gap. Instead, Mission champions would express a desire to do an IE, and DRG Center staff and academic partners would review programming for IE opportunities.[25] Given the myriad evidence gaps, this was not an unreasonable approach and it was responsive to Mission demand. Nonetheless, recognizing that a single IE has limited external validity (generalizability to multiple contexts) due to variation in settings, treatments, outcomes, units, and time, three USAID interviewees expressed disappointment that the DRG Center has not yet included the IEs in more comprehensive evidence reviews to inform a common learning agenda.

Moreover, as detailed in the next section, there were some IEs that went forward because an IE methodology was thought to be possible rather than because it was needed or desirable. Several interviewees noted that, given the level of effort and costs of IEs, it is better to cancel an IE than move forward with something that is not going to be useful.

IEs did not always test an adequately robust theory of change:[26] In several cases, the IE tested an intervention with a very weak theory of change that unsurprisingly resulted in null effect (e.g., South Africa, DRC, Zimbabwe, Peru). This left stakeholders frustrated with a long, difficult, and costly IE that did not produce valuable information. Such IEs went forward for at least three related reasons. The first was driven by presumed best practices in conducting IEs. Conventional theory in the IE community posits that IEs should be conducted on discrete interventions. If interventions have multiple elements to them, then detection of a positive effect will leave scholars and policy-makers uncertain as to what element of the intervention caused the change. While this is a valid concern, in some cases it led to overly simplistic interventions, such as trainings, which practitioners stated from the outset would not work. Second, interventions were at times hamstrung to be made evaluable. For example, in South Africa, one treatment arm intended to address community stigma to visiting a rape crisis center was unable to conduct mass education campaigns for fear of spillover effects on control areas. As such, the IE ended up testing the very unlikely effect of a one-time community dialogue event (attracting maybe 40 women) on short-term rape crisis center utilization for the whole community. Third, key elements of an intervention could often not be implemented as planned or were not well suited to the context of the intervention. For example, in Cambodia, an internet-based solution to human trafficking was implemented in villages with low internet penetration.

---

[24] Jonathan Rodden and Erik Wibbels (2013) Responsiveness and Accountability in Local Governance and Service Delivery: An Agenda for USAID Program Design and Evaluation. USAID.

[25] Prior to revisions to the ADS in 2020, Missions were to conduct an IE for each development objective (DO) and they are still required to conduct IEs of pilot initiatives, creating an incentive to conduct IEs irrespective of learning goals. Nonetheless, it is not clear if this was a significant factor in driving IEs. Most Missions did not follow the requirement as evidenced by the large gap between the number of DOs and pilots and the number of IEs. It is more likely that the requirement provided regulatory support for IE champions within Missions.

[26] A theory of change could be weak either because: (1) it follows a logical sequence but is not powerful enough to generate the expected results; or (2) the logical sequence itself is missing one or more key elements.

By contrast, in other cases, PIs prioritized a robust theory of change, leading to a stronger intervention (e.g., Malawi, Ghana, Uganda). In Uganda for example, PIs designed an IE of an SMS platform that allowed residents to make service delivery complaints to local officials. The planned platform, however, only allowed for one-way communication, and the PIs arranged for collaboration with other entities to implement a platform with two-way communication. Furthermore, the PIs realized that the intervention was not going to be adequately effective because not enough people would be enrolled in the text messaging platform. The IP lacked the budget to undertake more intensive registration efforts, so IE funds were repurposed to register residents in the platform. These same PIs recommended that a different IE of a local resource mobilization project not go forward because the proposed intervention did not adequately address the incentives for low tax payment.

In these various cases, a better-defined evaluation purpose may have helped to assess when more nimble formative evaluations were appropriate, even if a robust theory of change was not present, versus summative evaluations that required a robust theory of change. USAID generally, and DRG specifically, has thought of IEs from a summative perspective, but arguably most of its IE activities would be more formative. Thus, there has likely been a substantial mismatch between the purpose and the conduct of DRG IEs.

The role of academic PIs should match the goal of the IE: Academic PIs are generally regarded as a core pillar of the DRG Center's IE approach. PIs often played a central role in not just designing the evaluation but also the intervention. In Ghana, for example, the academic PI helped design an intervention that randomized Ghana's local governments into a treatment group receiving a top-down government audit conducted by the Ghana Audit Service and a bottom-up civil society scorecard campaign conducted by CSOs. The evaluation was regarded by the DRG Center as a flagship evaluation, testing two different theories about what works in creating local government accountability.

This PI-centric approach had both strengths and weaknesses. On the positive side, interviewees generally recognized the value addition of PIs. Interviewees noted that the PIs 1) bring theory and evidence to aid in developing approaches to testing, 2) have methodological expertise, 3) tend to be highly motivated, 4) are far less subjected to turnover than other IE stakeholders, 5) often do much of their work pro bono and even bring in complementary resources, and 6) bolster the independence of the IE.

On the negative side, academic PIs are not without their critics or limitations. Several interviewees highlighted the importance of personality as a key factor in success or failure. In the less successful cases, several IPs questioned why their expertise was subordinated in DRG Center IEs when they often had intimate contextual knowledge, subject matter expertise, and implementation expertise. As one IP respondent complained, "It was clear that the opinion of the implementers doesn't form a part of the evaluation." In fact, the survey suggests that PI-developed designs might not have been adequately informed by contextual knowledge. Only 29 percent of USAID and IP mini-survey respondents "strongly agreed" or "agreed" that evaluation designs matched the realities on the ground, as compared to 92 percent of evaluators (see question B4 in Annex 3, Survey Results).

Interviews suggest that the role of academics should depend on the purpose of the IE. If the goal of the IE is to test an existing approach that has shown promise but never been rigorously tested—a summative, confirmatory evaluation—then academics' role should focus on the evaluation design rather than the intervention design. This was the case in Cambodia, where NDI had been conducting community dialogues for many years but there was some uncertainty about whether it was working. It was also the

case in the youth violence prevention sector in the Caribbean, where an IP had developed a family counseling-based approach that appeared promising based on internal evaluations and was being exported to other contexts, but had not been tested with an IE design.

By contrast, in the case of Bangladesh, the Mission wanted to address a somewhat new threat of violent extremism but did not know the best way to proceed. PIs conducted a lengthy evidence review and identified a particular model as the most promising. This was subsequently piloted and tested and showed promising results—an exemplary model of an innovative, formative evaluation. As such, in cases where USAID and IPs do not know what will work and the goal is to innovate, there is a strong rationale for PI involvement in activity design.

Taken together, this suggests a key role for academics during formative, rather than summative, assessments. Even in the latter case, however, the implementers should still welcome critical feedback on potential weaknesses in their theory of change, as occurred in the Ugandan case described above. Moreover, there appears to be a broader lesson about the importance of a more collaborative process between the evaluation and implementation sides, each bringing a different set of skills and knowledge.

Another common challenge was that academic PIs operate on slower timelines than USAID and IPs, leading to complaints about long, drawn-out design processes (Uganda, Bangladesh) that undermined Mission and IP planning and resulted in long delays in producing final reports (Georgia, Mozambique, Bangladesh). Pairing PIs strategically offered an effective way to limit weaknesses. In some cases, pairing a more junior Assistant Professor or PhD student with more time and availability to travel with a more senior, experienced academic (as occurred in Bangladesh, the Caribbean, and Uganda) appeared to be a good strategy. In Tanzania, the PI team paired a methodologist and a Tanzania politics expert who spoke Swahili. In Ghana, there was an attempt to pair the US-based PI with a local Ghanaian PI to ensure adequate contextual knowledge. In the Caribbean, where frequent coordination with the IP was required, staff of the evaluation contractor carried out many PI functions, suggesting the possibility of an academic PI paired with a LP-based PI more knowledgeable about USAID programming.

IEs CONFRONTED SEVERAL IMPLEMENTATION CHALLENGES, BUT MANY WERE A PRODUCT OF INADEQUATE IP BUY-IN, INADEQUATE IP INPUT, AND CONFLICT BETWEEN EVALUATORS AND IMPLEMENTERS.

Survey respondents across the evaluations reported confronting numerous challenges in the course of the evaluation, and a majority of respondents reported challenges with both randomization and measurement. These occurred for a variety of reasons, many of them outside of the control of the evaluation and intervention. Weather-related problems had impacts on data collection and implementation in Malawi and Burkina Faso. Security concerns complicated data collection in Niger, the Cambodia Countering Trafficking in Persons (C-TIP) IE, the DRC gender-based violence IE, and in Mali. COVID-19 also had impacts on interventions in Tanzania and Bangladesh. Elections created timing challenges in Malawi and Ghana. Some challenges were the product of bad luck, or perhaps other data irregularities. In Malawi, despite a reasonable randomization process, treatment marketplaces had greater baseline tax revenue, the core outcome indicator, than control marketplaces. Other challenges emerged from human error; data collection firms performed poorly in Peru and Uganda, and data had to be thrown out. Indeed, data collection challenges were the single biggest problem cited by respondents in the survey.

In many cases, however, the challenges were the product of disagreements between evaluators and implementers or tensions between them. Although in most cases IPs respected random assignment, in Niger, Liberia, and Mali, respondents reported that the IP carried out some treatment in control areas. A limited number of units to randomize often meant that implementation had to be conducted inefficiently, on a national scale, in remote locations, or in locations (or to units) that the IP would never have selected as a priority for activity benefits. This was the case in South Africa, Cambodia C-TIP, Malawi, Tanzania, and in Paraguay. In Paraguay, for example, the project involved working with municipalities, and the implementer felt municipality buy-in was a precondition to successful outcomes. In other words, from the IP perspective, random selection of municipalities was not expected to work.

Many non-IP interviewees noted a tendency towards IP defensiveness and resistance to the IE that delayed the design process and undermined IE implementation and use. While the DRG Center tried to frame the IEs as about learning (i.e., as formative evaluations), IPs tended to view them as summative evaluations of their performance.

One group of interviewees referred to the DRG Center's model as a "forced marriage model" between implementers and evaluators, and interviews and surveys revealed considerable tension between these groups across both cooperative agreements and contracts. Several evaluation interviewees made comments along the lines of, "The IPs opposed the IE from the very beginning," while several IP interviewees made comments like, "The IE was measuring an activity that was completely ill-suited for [its] purpose." Despite considerable effort to build Mission buy-in, the DRG Center's approach lacked a corresponding effort to ensure IP buy-in and input. Given the fundamental role played by IPs, this was viewed by several interviewees as major and preventable shortcoming.

MOVING FROM A CONFLICTIVE TO A COOPERATIVE RELATIONSHIP BETWEEN IMPLEMENTERS AND EVALUATORS

The DRG Center IE experience produced a number of lessons learned in obtaining IP buy-in and fostering cooperation between evaluators and IPs. The Center needs strategies at the outset to move from a conflictive to a cooperative relationship between implementers and evaluators.

The importance of clear solicitations: As noted above, language was often included in solicitation documents that there would be an IE. However, in many cases this language was insufficient and did not clarify the implications of the IE (e.g., Caribbean, Malawi, Tanzania). For example, solicitation language might note that there would be an IE and the IP should collaborate but fail to note the implications for the selection of beneficiaries, time and budget expectations, implementation fidelity requirements, and information and M&E sharing expectations. In these cases, IPs failed to adequately budget the time and money for the IE or understand the requirements in terms of randomization, implementation fidelity, reporting, and communication.

In a handful of cases, the solicitation was more detailed and the IPs felt adequately informed. This was particularly the case in what one interviewee referred to as "grafting IEs." These were cases where a prime IP would subcontract with a local organization to carry out the intervention to be evaluated (Paraguay, Bangladesh). In other words, the evaluated intervention was grafted on to an existing award. In the Paraguay case, the PIs conducted a workshop on implementing an IE and its implications with interested bidders prior to the bid. In both of these cases, PIs helped develop the bidding language and sat

on the technical evaluation committee selecting the IP. This worked well for formative evaluations where the PIs were designing both the intervention and evaluation.

The need for intensive post-contracting IP stakeholder buy-in efforts: In several cases, onboarding of the IP led to a contracted period of frustrating negotiations between the IP and the evaluators, and the DRG Center's experience suggests that intensive efforts are also needed after procurement. In several cases (e.g., Bangladesh, Paraguay, and Cambodia-NDI), PIs led multi-day workshops with the IP designed to teach counterparts the basics of IEs and create excitement over the IE, much like was done with the IE clinics. Moreover, the workshops allowed for two-way communication, whereby PIs learned contextual and programmatic objectives to inform selection, randomization, intervention elements, and measurement. This resulted in constructive IP input in the design. In Bangladesh, the IP suggested a change in which the intervention was implemented and in Cambodia-NDI, the workshop allowed the evaluation team and IP to determine a randomization strategy. Such a workshop mirrors the utility of the IE clinics.

It is worth noting that none of the case study IEs, with the possible exception of the Caribbean, produced a formally signed Memorandum of Understanding (MOU) or a less formal document that clearly laid out expectations and responsibilities for each stakeholder as recommended by USAID's Bureau of Policy, Planning, and Learning (PPL) (see PPL sample). In some cases (e.g., Malawi, Georgia), IPs reported not having a document that even detailed the design.

The problem of the evaluation-implementation firewall: Despite these efforts, the forced marriage model still resulted in a clear divide between evaluators and implementers. Conventional IE wisdom in USAID, likely rooted in a summative evaluation framework, highlights the importance of evaluator independence; however, this frequently led to conflict. Several interviewees reported contentious meetings with voices raised and prolonged periods to come to agreement on key decisions. In several instances, USAID staff had to spend excessive time mediating disputes or demanding contractual or design compliance by one side or the other.

In response, several interviewees cited the need to deemphasize the firewall between implementer and evaluator in order to make the metaphorical marriage successful. In fact, J-PAL, IPA, and the Department of Labor—Bureau of International Labor Affairs (ILAB) apply a different model that does not include such a rigid firewall, and they have successfully maintained objectivity while reducing evaluator-implementer conflict. Under this model, PIs find governments or IPs willing to collaborate on an IE and then apply for funding. At the World Bank, which has conducted perhaps more evaluations than any other development actor, the evaluators and implementers (for them, the country government and ministries) work hand-in-hand from the very beginning. One USAID interviewee commented, "We need to get over the myth of complete independence of the evaluation." As such, in a more recent IE, this individual noted that IP MEL representatives are involved in all IE meetings as part of the same evaluation team. This relaxing of the fire wall does not have to come at a cost of objectivity. Interviewees note that objectivity can still be obtained through 1) academic professional norms, 2) increased transparency, including pre-registration of how data will be analyzed, 3) the inclusion of an objectivity risk mitigation strategy as part of the design

and design discussions, and 4) the absence of a fiduciary relationship between the PI and the IP. This is entirely consistent with ADS requirements on evaluation independence (ADS 201.3.6.6).[27]

In fact, several evaluations illustrate a need for closer cooperation between implementers and evaluators in data collection. In the Caribbean, the baseline data was derived from a risk assessment tool developed by the implementer, and the IP and evaluator needed to work together to field the assessment. In Malawi, the IE was under-resourced, and relied on the IP to collect administrative data on local tax collection as part of their efforts to improve local monitoring capacity, a task that turned out to be far more labor-intensive and challenging than expected. In Haiti, the evaluation team hired the same lawyers working with the IP to collect data in the prison, as only they had the detailed knowledge and contacts to obtain the required information. Failure to coordinate properly led to delays (Caribbean, Malawi) and failed data collection (Tanzania). Earlier and better coordination, and indeed preemptive coordination, would help immensely when engagement is likely to happen anyway in many cases.

A need to ensure that the intervention was ready to be tested: The majority of IEs were conducted at the beginning of an intervention or contract period. IPs reported challenges with this model, particularly if they were tasked with a complex intervention involving multiple interventions (Malawi, Uganda, Tanzania, Caribbean). One interviewee noted that the first year of an intervention is often spent figuring things out: onboarding staff, engaging stakeholders, developing relationships, testing assumptions, and laying the groundwork for future activities. Chiefs of party and deputy chiefs of party were often required to spend a disproportionate amount of their time on the evaluated component of their intervention to the neglect of other priorities, generating frustration. Furthermore, IPs might be new to the content (Tanzania, Bangladesh) or the context (Caribbean), and they might need time to be able to effectively work out implementation challenges and adapt their approach. In the Caribbean, for example, the IP was replicating an approach they had implemented in Central America, but it was new to the Caribbean, and the IP did not have the necessary relationships to hit the ground running. The DRG Center had a filtration process to identify potential IEs and filter out poor candidates. However, unlike MCC and other donors, the DRG Center does not require a formal evaluability assessment to ensure the intervention is ready for an evaluation.

The evaluation team observed two solutions to this problem. In the Haiti IE, the intervention was tested in its final year after the implementer had the knowledge, experience, and relationships to be successful in a challenging operating environment. In Bangladesh, a training-based intervention was piloted and adapted prior to being scaled up for the evaluation. A similar pilot-based approach would have benefited the Malawi IE, where the intervention turned out to be complex and challenging for the IP to implement across multiple remote locations. This approach addresses another priority: adaptation. USAID currently promotes implementation adaptation; however, this can complicate an IE, which generally requires consistent application. By conducting an initial pilot, the IP and evaluators had an opportunity to adapt the implementation approach before it was tested. In any event, a better-defined evaluation purpose, whether formative or summative, with more specifics as appropriate, would help shape expectations about what activities were required, how early, and how defined.

The need for strategies during IE implementation for improved communication and coordination: Rather than a forced marriage, one interviewee argued that a better analogy was a

[27] Hageboeck et al. Impact Evaluation: Critical Challenges/Promising Solutions arrive at a similar conclusion and recommend joint design development through a post-award workshop and joint scoping trip.

dysfunctional family in need of a family counselor. In fact, the implementer and evaluator are not the only IE stakeholders and these groups can be further subdivided. Most IEs require communication and coordination across PIs, an evaluation contractor, a data collection firm, the DRG Center, USAID Mission program and technical offices, IP project teams, IP headquarters, and IP subcontractors or government partners. Several interviewees noted the different interests and incentives among these groups: PIs looking for publications and tenure, contractors looking to minimize costs, and IP field staff trying to meet output targets. Any one of these stakeholders can undermine the success of an IE, and one can point to missteps by all of these stakeholders. Several interviewees expressed a preference for reducing the number and type of stakeholders.

The need for a communication strategy: In many cases, these natural divisions were exacerbated by inadequate communication and coordination. With stakeholders all juggling multiple priorities, there were long periods of time in several IEs without any communication between stakeholders. This led to a number of complications. In the Malawi revenue mobilization IE, for example, the IP undertook a radio campaign promoting tax compliance that could have contaminated the control group without mentioning it to the PIs. On the other side, the IP in Malawi did not see the evaluation report until it was finalized and on the Development Experience Clearinghouse (DEC).

A common challenge related to implementation fidelity, consistency, and reporting. PIs frequently expected intervention standardization, monitoring systems that tracked implementation quality, and shared reporting on implementation. By contract, several interventions were implemented by multiple sub-grantees or sub-contractors; IPs were far less concerned with standardization and they were hesitant to share information. In Ghana, the evaluation team reported needing to use their own time and resources to develop intervention monitoring.

As discussed above, clear solicitations, a post-contract workshop, documentation of responsibilities, and efforts to create a team-based approach can help improve coordination and communication among multiple stakeholders. Regularly scheduled meetings with follow-up emails, a process for documenting agreements, and a shared drive for sharing information all appears essential to coordination across so many stakeholders. Unfortunately, even when these measures did occur, they often emerged reactively after problems had occurred rather than proactively.

USAID plays a key facilitation role: If an IE is a dysfunctional family, then several interviewees felt that USAID staff needed to play the role of family counselor. In many cases, USAID Mission staff and DRG Center staff played a key role in mediating between IPs and evaluators. Sixty-six percent of survey respondents agreed or strongly agreed that USAID effectively balanced the needs of the intervention and the needs of the evaluation, although IP respondents were somewhat less in agreement than other stakeholders (see C2; Annex 3, Survey Results). USAID did face two major limitations in playing this role. First, at the DRG Center, staffing fluctuated and there was often inadequate bandwidth to engage substantively on evaluation challenges. Second, at the Mission level, many IEs were originally championed by Foreign Service Officers (FSOs) that moved on to other posts in the course of the IE. In a handful of cases, DRG Center staff attempted to serve as PIs for the evaluation. This did not generally work well, as DRG Center staff lacked the bandwidth to play such a role. In the worst-case scenario, a DRG Center PI left their position after baseline data collection without transferring a design document or important details on the baseline (Georgia).

Evaluation teams often needed an in-country presence: Primarily due to budget limitations, most IEs entailed limited evaluation team presence in the field, which was widely seen as undermining IE effectiveness and exacerbating communication and coordination challenges. Many PIs only made a handful of visits after the original scoping trip and only a few of the IEs included staff based in the country, a requirement for IEs conducted by MCC, IPA, and J-PAL. By contrast, in the Bangladesh IE, one of the PIs made five trips over the course of two years, which was cited as a key factor in the IE's successful implementation. In-country evaluation team members in Uganda and Mozambique were considered key to the success of those two IEs. Such staff can maintain communication with the IP, oversee data collection, and conduct site visits to observe implementation.

LPs' value added was limited when they were a financial passthrough: Interviewees expressed mixed views on the role of the DRG Center's LPs. For some interviewees, their role was absolutely indispensable, freeing PIs to focus on evaluation substance and ensuring coordination with data collection firms, IPs, and USAID. One interviewee noted that doing the work without them would be "delusional." Moreover, in theory at least, the LPs should have experience across multiple evaluations that they could bring to bear on IEs and ensure that some of the practices identified here are taken up by less experienced PIs, Missions, and IPs. In several instances, however, interviewees questioned their value addition and criticized a lack of financial transparency with PI partners.

The LPs' value appeared more likely to be questioned when not actively involved as a part of the evaluation team (Uganda, Ghana, Tanzania, Malawi). In some cases, LPs' functions were limited to coordination functions and procurement of data collection partners; however, absent engagement in the day-to-day details of the evaluation, time in the field, or first-hand relationships with Missions or IPs, they were unable to perform their coordination responsibilities and were more likely to serve as just one more stakeholder. In cases where well-qualified mid- or senior-level evaluation contractor staff were engaged in field work (e.g., Bangladesh, Paraguay, Caribbean), their value addition was appreciated. In these cases, they brought a mix of local knowledge, methodological expertise, experience with USAID programming and processes, and experience across multiple DRG Center IEs.

CONCLUSION

Despite significant achievements, the DRG Center's IE program encountered several challenges, many of which were common across IEs. These generated a number of lessons learned to inform future IEs. First, the objective and intended use of the IEs was often not well defined. For example, it should be clear whether the goal of an IE is to help *determine a new USAID approach* to addressing a DRG program (i.e., a formative IE) or to *test USAID's existing approach* to addressing a DRG problem (i.e., a summative IE). The second set of challenges and lessons learned concern the decision to conduct an IE. Just because an IE was possible did not mean that one should have been carried out. For example, several IEs failed to test an intervention with an adequately robust theory of change. Furthermore, while academic PIs are generally regarded as a core strength of the DRG Center's IE approach, their role should match the goal of the IE, and it did not always make sense for the PIs to play a lead role in designing interventions. Third, DRG Center IEs confronted numerous implementation challenges, but many of them had their roots in inadequate IP buy-in, insufficient input from IPs, and conflict between implementers and evaluators. Fourth, as such, the DRG Center and its partners needed strategies at the outset to move from a conflictive to a cooperative relationship between evaluators and IPs. Effective practices included clear solicitation language, intensive post-contracting stakeholder engagement efforts, a weakening of the traditional firewall

between evaluators and implementers, and assurance that the intervention is ready to be tested prior to the initiation of the IE. Fifth, the DRG Center and its partners also needed strategies to improve coordination and communication across IE stakeholders during implementation, including agreement on communication and information sharing protocols, active DRG Center and Mission engagement and facilitation, an in-country presence for evaluation teams, and an active role for DRG Center LPs. Finally, many evaluations did not meet stakeholders' expectations. USAID and IP survey respondents were considerably less likely than evaluators and PIs to identify the benefits of IEs. In some cases, null results could not be explained, mixed results did not produce clear policy guidance, evaluations of weak interventions produced little value added, and there were limits to what an IE could test.

5. HOW HAVE USAID AND OTHERS USED THE IEs?

Question 4: Use: How has USAID (or others) used the IEs? Why were some evaluations more useful than others? How have findings been disseminated?

Utilization of lessons learned is a key objective for all IEs. The retrospective team examined four different types of utilization: *positive programmatic utilization*, *negative programmatic utilization, positive design utilization,* and *negative design utilization*. Positive programming utilization occurs when a later program or activity is tied into earlier lessons learned or IE findings. Negative programmatic utilization occurs when future programs are not created due to learnings from an IE and is more challenging to measure than positive programmatic utilization. Similarly, positive design utilization occurs when activity design is changed based on IE findings, and negative design utilization occurs when activity designs are not implemented due to IE findings.

This retrospective found little demonstrable evidence of positive utilization, and only hints of negative utilization on the programmatic side, though there is more evidence of each type for design utilization. The utilization that the team found varied widely across IEs.

In response to Question 2, the team found that the DRG Center's IE program produced some important and valuable findings; however, there are several steps in getting from findings to use. A prerequisite to use is that evidence needs to lead to learning, or at least an interest in learning.[28] While IEs might produce evidence, if that evidence is not clearly communicated, disseminated to the right people, and discussed, or if it is simply dismissed by stakeholders (e.g., due to defensiveness, doubts about the methodology or quality of the evaluation, etc.), it is unlikely that the evidence will result in learning within USAID or among its IPs[29]. This section assesses how IEs were used, explores reasons why utilization was more successful in some cases than others, and considers the role that dissemination plays.

[28] From Evidence to Learning: Recommendations to Improve U.S. Foreign Assistance Evaluation (2017) The Lugar Center and Modernizing Foreign Assistance Network.

[29] Political economy factors also influence the utilization of IE results, and some programs may take place (or not take place) for political reasons, regardless of the evidence base.

HOW HAS USAID (OR OTHERS) USED THE IEs?

When USAID's Evaluation Policy[30] was created in 2011, it created a robust system to increase evaluation in USAID, including clear policy mandates, extensive guidance documentation, intensive training, and funding requirements. Nonetheless, neither the policy nor the ADS included a focus on how evaluation findings would be used. In 2016, the ADS was revised to require the production of a post-evaluation action plan, but anecdotal evidence suggests that this guidance is frequently not followed and the team did not find evidence of extensive use of post-evaluation action plans in this study, despite 15 being complete in or after 2016. Broader success stories of IE utilization include IE findings leading to national policy changes in Haiti and improvements to an intervention's theory of change in Uganda; however, in most evaluations, stakeholders were not able to identify direct changes resulting from the IE. This section explores the four types of utilization: usage in existing program design (positive and negative), usage in strategy and programming (positive and negative), as well as expansion/refinement of the DRG knowledge base.

IE INFLUENCE ON EXISTING PROGRAM DESIGN

Final IE findings were often available too late in the program cycle to inform decisions about future activity implementation, such as the decision to scale up. For example, in Bangladesh, the Mission originally hoped that the IE would occur in year one or two of the activity and inform programming in the final year; however, the results were not completed until after the activity was closed. In Uganda, the implementer had to start planning and budgeting its intervention scale-up long before the evaluation report was completed.

Nonetheless, in several cases, midline and endline findings, as well as discussions between the IP and evaluator, led to improvements in the USAID project/activity being evaluated. For example, in Uganda, the evaluation team convinced the implementers to use a better SMS platform that allowed for two-way communication. In the Caribbean, the evaluation team suggested a re-design of a youth risk assessment tool used by the program, which the IP adopted. In Ghana, the IP team expressed that findings of the IE were much more useful for improving their program implementation than the findings of a traditional three-week PE. They attributed this to the duration of the IE and the depth of knowledge about the program the evaluation team developed. IE evidence and recommendations helped build an audit structure for the government of Ghana that is still in use today. Similar stories of program adaptations informed by the IE can be found in 10 of 22 IEs with survey responses, and 32 percent of survey respondents agreed with the statement, "The IP used IE findings to make decisions about the program being evaluated." Table 3, below, shows the number of IEs that report various types of utilization as a percentage of 22 IEs with respondents.

TABLE 3. IE PERCEIVED UTILIZATION AMONG SURVEY RESPONDENTS

USAGE	PERCENTAGE (N)[31]
Informed the implementer's understanding of the development challenges	14
Informed USAID's understanding of the development challenges	13
Resulted in changes to the intervention being evaluated	10

[30] Evaluation: Learning From Experience USAID Evaluation Policy (2016). USAID.

[31] 22 IEs had at least one survey respondent

USAGE	PERCENTAGE (N)[31]
Informed USAID country or sector strategy	9
Informed guidance, training, or assessment material	9
Data was an input to the monitoring, evaluation, and learning (MEL) plan	8
Resulted in changes to future evaluations	8
Informed intervention design in other countries	5

IE INFLUENCE ON USAID STRATEGY AND FUTURE PROGRAMMING

According to the survey, nine of 22 IEs with survey responses informed USAID country or sector strategy, and five IEs informed USAID decisions about similar activities in other Missions. However, when asked an open-ended survey question of how the evaluation was used, only one respondent offered an example of positive utilization. The most important example of IE use occurred in Haiti, and is explained in the text box below. Interviews also produced few examples of influence on USAID strategy, and these were difficult to verify. One former DRG staff member noted that IE findings from Uganda influenced programming in Senegal, through a connection made by the DRG Center. In Ghana, a technical expert designing a new USAID accountability project in 2020 reviewed the IE but it does not appear this informed her decision-making. This may stem from a lack of awareness across Missions about what work is being done on particular topics, IE findings that are difficult to apply to other contexts (or lack guidance on how to do so), or IE recommendations that overlook USAID's wider strategic and programming goals. Addressing this gap in utilization is critical to increasing the value of IEs to the DRG Center.

IE FINDINGS LEAD TO NATIONAL POLICY IN HAITI

In Haiti, the IE found that nine months after the start of the IE period, legal assistance increased the proportion of detainees who were freed. The evaluation team recommended the creation of a public defender office and the adoption of a new criminal procedure code and a new penal code. These recommendations were used by the Mission to inform the design of a new USAID/Haiti judicial strengthening activity which started right after the end of the evaluated PROJUSTICE activity. Among other things, the new activity was specifically tasked to provide technical assistance to the Haitian Government and other key stakeholders for the implementation of the recommendations from the IE. As a result of that assistance in 2018, Haiti adopted a law creating the public defender's office, which would provide free legal aid to pretrial detainees. With support from the new project, the Haitian Government also adopted a new criminal procedure code and a penal code in 2020, and funds for the legal aid were included in the 2021 budget.

IE INFLUENCE ON THE KNOWLEDGE BASE

The most common utilization of IEs cited by survey respondents was that the IE informed broader USAID and IP understanding of development challenges. One academic PI interviewee even contended that USAID strategy and programming should never be based on the results of a single IE, but rather on a learning agenda that emerges over time and across programs. IEs have also led to numerous academic papers, journal articles, and conference presentations, though the exact number is not documented. These

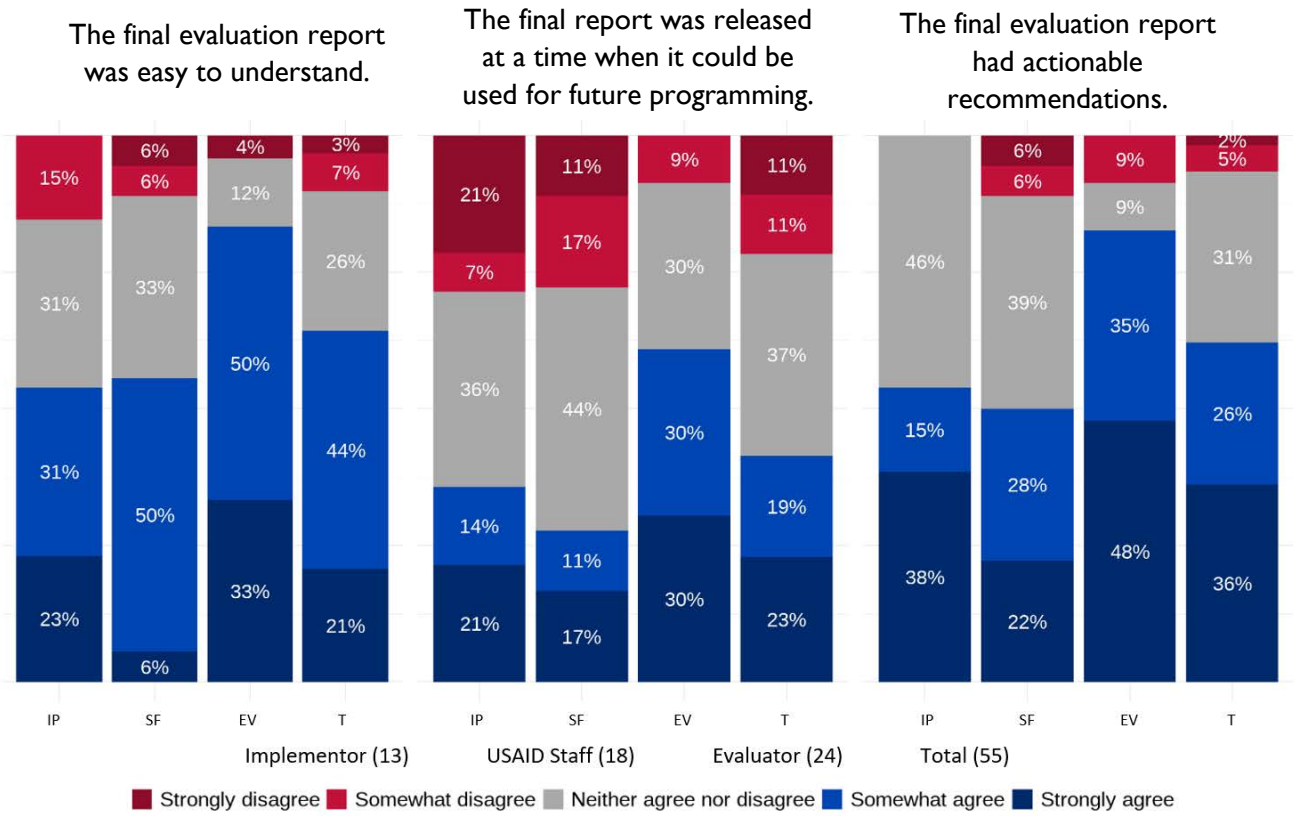
contributions to the global literature present an opportunity for USAID to position itself as a global thought leader in DRG research, and a pioneer in how to measure DRG outcomes. These goals were part of the original vision for IEs at the DRG Center, though DRG past and present staff have mixed opinions about whether providing research as a public good is an appropriate priority. For many interviewees, contributing to the public good of evidence alone is not sufficient usage.

WHY WERE SOME EVALUATIONS MORE USEFUL THAN OTHERS?

There are several factors that might explain evaluation use, and some have already been mentioned. Timeliness was clearly important in the cases of Bangladesh and Uganda. As shown in Figure 7, only a minority of USAID and IP survey respondents felt that the final evaluation report was released at a time when it could be used to inform future programming. In Ghana, the Mission credited detailed data and knowledge generated over time for it's high perceived utilization of the IE. Other factors explored in this section include the nature of the findings, dissemination strategies, and commitment to use.

Perhaps unsurprisingly, survey respondents and interviewees associated with IEs with positive results report greater utilization than those associated with IEs with negative, mixed, or null results. Evaluations that scored lowest on a measure of whether the final evaluation could be used by future programming were evaluations with null impacts (Zambia, Peru) or mixed results (Nepal, DRC), which suggests IPs and Missions are less willing to let negative or null results influence decisions about programming than they are positive results.

Figure 7. Evaluation Report utilization

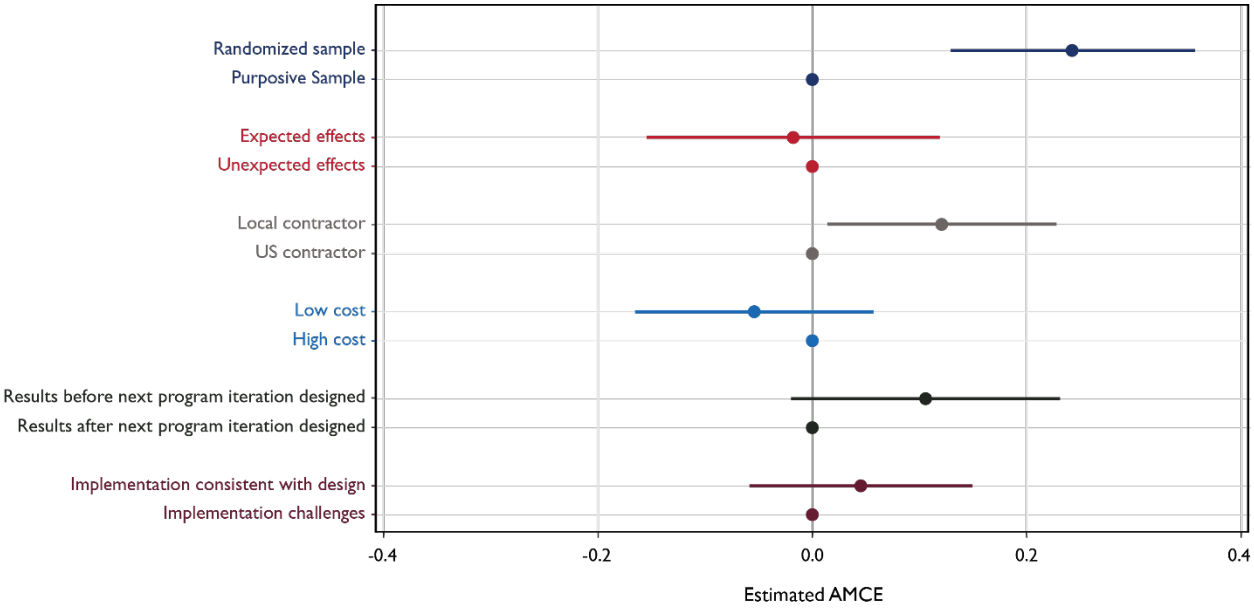


RESULTS OF A CONJOINT EXPERIMENT

Because the 27 IEs lacked variation in some respects, the team conducted a conjoint experiment as part of the quantitative survey, and considered individual components of IEs and what respondents found most preferable. The conjoint offers some useful lessons for why some types of IEs would be more useful than others. The team considered six components of IEs, including whether: 1) the intervention was randomized or purposively assigned; 2) the results were null or confirmed the theory of change; 3) the IP was local or international; 4) the costs of the evaluation were low or high (two percent or four percent of total project costs, respectively); 5) the results were released before or after the following program cycle; and 6) the implementation of the activities occurred with challenges or as planned. After seeing profiles of two IEs side by side, respondents were asked, "Which of the IEs would contribute most to USAID usage of the evaluation results?" Aggregating these comparisons allowed us to isolate the marginal effects of each individual component.

The results of the conjoint experiment revealed several key findings. See Figure 8 for overall results. Although disaggregating by stakeholder type reduces statistical power substantially, the team examined how the results may have differed by type. First, evaluations in which the intervention is randomized to intended beneficiaries are preferred (statistically) to evaluations in which assignment is purposive. Although all stakeholder types expressed a positive preference, the statistical significance of the result appears, perhaps unsurprisingly, to be driven primarily by PIs and evaluators. Second, evaluations with local evaluation partners are statistically preferred to those with only international evaluation partners. When considering stakeholder type, it appears that the preference for local evaluation partners is driven by USAID staff and IPs. Third, there is no overall preference for lower or higher cost evaluations but, when disaggregating by stakeholder type, academic PIs and evaluator LPs are less likely to prefer lower-cost evaluations. None of the results for confirmatory/null results, timing of the release of results, or presence/absence of implementation challenges were significantly different in either the aggregated or disaggregated analyses. In sum, the conjoint experiment gives some insight into why some IEs would be used more than others and, echoing a common theme, each stakeholder's constraints and incentives appear to play an important role.

Figure 8: Conjoint Experiment Analysis



HOW HAVE FINDINGS BEEN DISSEMINATED?

Success stories include IE findings leading to national policy changes in Haiti and improving program implementation in the Caribbean. On the opposite end of the spectrum, some IEs, such as the IE of a media program in Tanzania, have had very little demonstrated use. Table 4 shows the number of IEs that report various types of utilization. Numbers are reported as a consolidation of responses across stakeholder groups and may overestimate utilization if one stakeholder believes another stakeholder used the findings in a way they did not.

TABLE 4. IE UTILIZATION

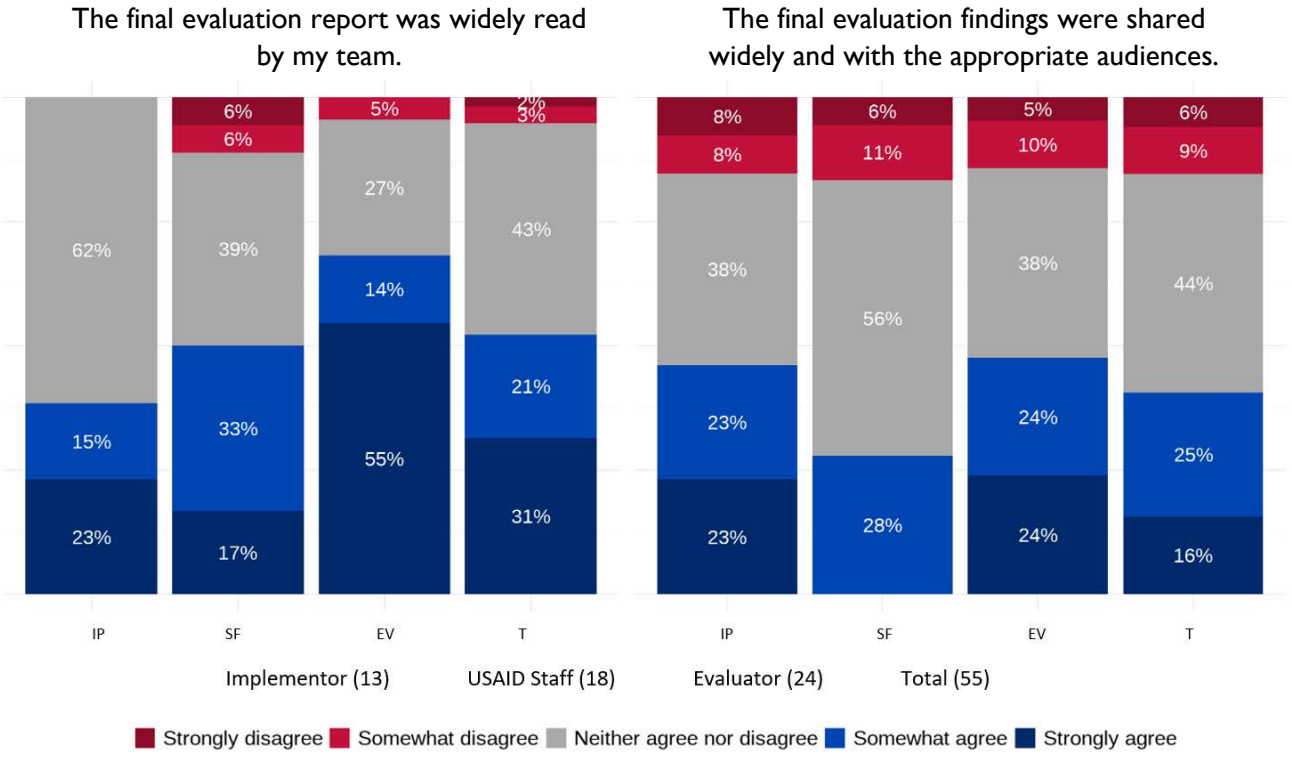
USAGE	N
Informed the implementer's understanding of the development challenges	14
Informed USAID's understanding of the development challenge	13
Resulted in changes to the intervention being evaluated	10
Informed USAID country or sector strategy	9
Informed guidance, training, or assessment material	9
Data was an input to the MEL plan	8
Resulted in changes to future evaluations	8
Informed intervention design in other countries	5

Another factor that appears to explain variation in utilization is the extent of dissemination and the presentation of the report itself. As part of the typical IE implementation, evaluators and PIs prepare reports after each round of data collection, including a larger findings report, which is circulated to USAID and often, but not always, the IP. Less often, the report is circulated to other stakeholders, such as host country governments. Forty percent of survey respondents agreed or strongly agreed that the IE findings

were "widely shared." Most evaluation reports were uploaded to the DEC, although some of the first-generation reports could not be found there. Many were also complemented with a two-page summary of evaluation findings in a more reader-friendly format, with images and visualizations.

As shown in Figure 9, survey responses suggest that evaluation reports were not widely read or shared. A minority of IP respondents felt that the evaluation report was widely read within their team, and only 28 percent of USAID respondents felt that evaluation reports were widely shared.

Figure 9. Audiences For Evaluation Findings



In many cases, a dissemination event occurred at USAID headquarters and/or in the host country (e.g., Ghana, Uganda), but this was not a standard practice and many IEs only had limited dissemination. Uganda, discussed in detail below, stood out as a model of dissemination; however, in both the Ghanaian and Ugandan cases, PIs used their own university-based research funds to subsidize the cost of dissemination. In some cases, where the IE report was poorly received by the IP or the Mission (e.g., Malawi, South Africa), there appeared to be a desire to simply move on with little appetite for dissemination. As one respondent put it, "There's usually less excitement about a report with null results."

The DRG Center's main strategy to disseminate to a wider USAID audience was through summaries in its monthly newsletter. The DRG Center also attempted to make some more visually appealing infographics and summary documentation for a few IEs. Findings reports and datasets are also made available on the Development Experience Clearinghouse and Development Data Library. Given this wide sharing, it is likely that findings have been used in more ways, but the Retrospective Team had no way to measure or track this utilization. Generally speaking, however, survey responses suggested that the DRG Center, Missions, and partners could have done much more dissemination.

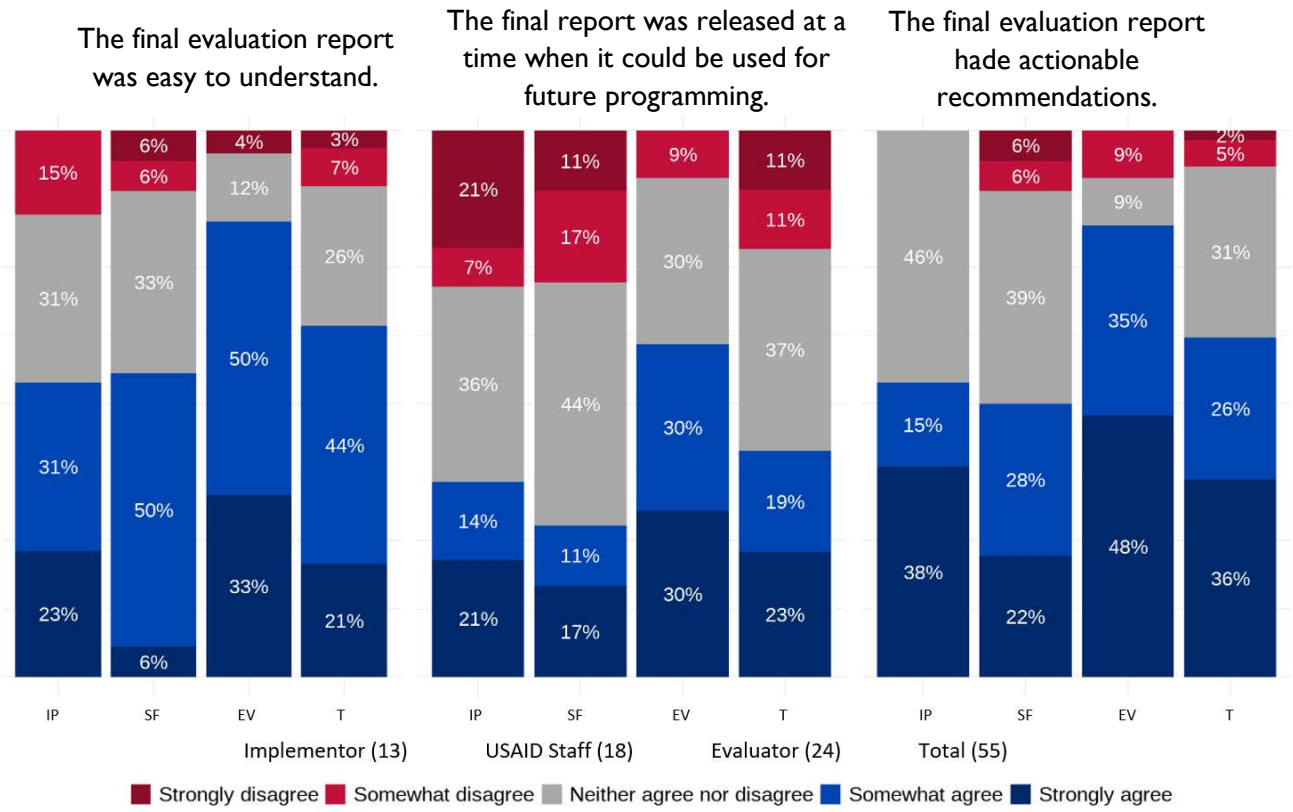
More innovative strategies for sharing results, such as through media (blogs, podcasts, editorials) were uncommon or nonexistent. Some survey respondents cited lack of time and lack of budget in the contract as limitations to more extensive or non-traditional dissemination strategies. Post-evaluation action plans, required under revisions to the ADS in 2016 (201.3.6.10.a.), are beginning to gain traction and have the potential to play an important role in ensuring evaluation findings make the jump from being shared to being used. While some survey respondents reported the use of post-evaluation action plans, none of the case studies produced such a plan.

DISSEMINATION AND UTILIZATION OF GAPP IE FINDINGS IN UGANDA

At the conclusion of the GAPP IE, the team incorporated a suite of dissemination activities. In addition to distributing the findings report and producing two evaluation briefs, USAID/Uganda, the research team, the IP RTI, and local government officials in two districts held two dissemination workshops with program beneficiaries to share and validate findings. Citizens and leaders had the opportunity to talk about what the research showed and, subsequently, district officials implemented IE recommendations, including creating case logs for requests received through the SMS system.

Dissemination and utilization often begin with the IE findings report, a highly technical, often lengthy, document produced by the PIs and evaluators. By design, the findings report includes regression analysis, power and balance calculations, and recommendations that tend to orient more to the academy than to practitioners. Figure 10 below shows how the final evaluation report was *perceived* by stakeholders. While evaluators and PIs generally agreed or strongly agreed that evaluation reports were easy to understand, USAID staff and implementers were far less likely to agree. In addition, only around half of USAID and IP respondents agreed or strongly agreed that the IEs included actionable recommendations. The document review found that many reports were more likely to offer recommendations for further research than for programming. Moreover, interviews suggest that conclusions were often more tentative than IPs and USAID Mission staff may have expected. The figures reveal a clear difference of perception across the evaluator and practitioner division. One IP said of the findings, "They were not so conclusive for us to draw a lot on them. We were looking to understand more about what leads to increased revenue collection and didn't find much direct impact on this."

Figure 10. Perceptions of the Evaluation Report



In summary, there are several important examples of how DRG Center IEs have been used. The most salient of these was in Haiti, where the IE helped justify legal reforms and the government taking over the legal defense of pretrial detainees. While we do find evidence of IEs informing existing projects, future projects, strategies, and general knowledge, there is considerable variation in IE usefulness. Several factors help explain this variation. Survey and case study evidence show that IE reports are often produced too late to inform decision making. In addition, although there are good examples of dissemination, on the whole the survey suggests that reports were not widely read or distributed. Furthermore, while evaluators generally reported that reports were easy to read and contained actionable recommendations, IPs and USAID survey respondents were far less likely to agree. Finally, although post-evaluation action plans are a USAID requirement, they were the exception rather than the norm.

6. WHAT SHOULD THE DRG CENTER'S APPROACH TO IEs BE GOING FORWARD?

Question 5: Recommendation: What should be the DRG Center's approach to IEs moving forward? Under what conditions are they most effective and useful? How could the DRG Center better support Missions and others in the utilization of IE findings/recommendations?

It is clear that the DRG Center's IEs have made a positive difference in some USAID programming, and yet, given many challenges, it is difficult to consider the program an unmitigated success. How should the DRG Center respond? Most survey and interview respondents, across all stakeholder types, feel that the

goal of testing the effectiveness of DRG interventions is too important to be abandoned and that the DRG Center should re-initiate its program and learn from the errors of past evaluations. By contrast, a handful of interviewees felt that the benefits of the DRG IE program did not outweigh the costs, that the decision to scale back the program in 2019 was justified, and that the new status quo with minimal IEs should continue.

The evaluation team sees three major options for the DRG Center moving forward.

- **Option 1:** Maintain the relatively new status quo of supporting Missions only when they want to do IEs and have the technical capacity. This excludes DRG Center promotion of IEs, co-funding, intensive Evidence & Learning Team technical support, and the IE clinics.
- **Option 2**: Build from the previous IE program, leveraging the optimism of most stakeholders, as well as the key advances, but making changes to address the concerns identified in this report. This includes DRG Center IE promotion, co-funding, and intensive Evidence and Learning Team technical support, and potentially the IE clinics.
- **Option 3:** Shift to a grant-making approach (as done by ILAB and J-PAL), whereby entrepreneurial PIs find willing, and often also entrepreneurial, IPs and then apply for funding.

Based on an aggregation and synthesis of input from interviewees and survey respondents and the desk review, the team recommends Option 2 over Options 1 and 3. The evaluation team believes that Option 2 is preferable to 1 on several grounds:

- First, it is important to note that USAID policy still requires IEs and USAID in general, and DRG in particular, are currently not compliant with the spirit of the policy, if not the letter.
- Given the many challenges of conducting IEs, it is not recommended for Missions to try to conduct IEs on their own. Absent experience and expertise, they are likely to fall into the many pitfalls identified in this study. Any multi-stakeholder coordination challenges, including those associated with the inclusion of professional evaluators and academic PIs, are eclipsed by the greater problem of carrying out low-quality IEs or, worse, no IEs at all. In fact, this was one of the lessons learned from the first generation of DRG Center IEs. Moreover, as discussed in the Question 1 response, Missions and IPs do not have an independent incentive to undertake IEs.
- The DRG Center, however, does have the incentive and mandate, and it is well positioned to promote and support the development of evidence to inform DRG programming. In fact, the DRG Center could be doing much more to improve both the conduct of as well as the learning potential from IEs. In addition to this report, in recent years other USAID operating units have generated similar lessons learned and guidance documents that can be adapted and built on.[32]

- Finally, most surveyed stakeholders felt that IEs should continue or be increased. Only three survey respondents felt that the DRG Center should cease doing IEs, and pluralities across stakeholder groups, including IPs, felt the DRG Center's IE program should continue (see Figure F5, Annex 3, Survey Results). Indeed, there seems to be broad support for IEs to the extent they can minimize friction and help an array of stakeholders perform their responsibilities better.

[32] See supporting guidance in Hageboeck et al. Impact Evaluation: Critical Challenges/Promising Solutions and Velez. Assessing the Quality of Impact Evaluations at USAID.

Option 3 is a valid option that could be considered. Still, Option 3 is unlikely to generate integrative and cumulative knowledge for the agency, or for the DRG Center in particular, because it relies on something of an opportunistic and entrepreneurial mindset, which has consequential implications for broader learning. Most notably, without sustained attention from the DRG Center to coordinate the integration and cumulation of evaluation evidence, Option 3 comes with severe selection problems that make learning spotty and biased towards convenience. A grantmaking approach under Option 3 could play a supporting role, and the DRG Center could strategically offer grants to incentivize some types of IEs that would otherwise not emerge or be carried out.

To proceed with Option 2, the team outlines several key recommendations. Consistent with the challenges identified earlier, it is critical to appropriately identify the purpose of the evaluation, whether that be formative or summative, across all recommendations; the purpose should shape each of the specific recommendations.

AN EVALUABILITY ASSESSMENT SHOULD INFORM WHETHER TO CONDUCT AN IE AND FOR WHAT PURPOSE

Given that most IEs moved ahead because of Mission demand, or simply because they were possible, evaluability assessments should either be required or strongly encouraged.[33] If USAID is considering an IE, it should first carry out an evaluability assessment during project design that confirms that a project is suitable for a quality IE. A project's suitability depends on a number of factors including a sufficiently strong theory of change, the feasibility of evaluation needs such as adequate units for purposes of randomization, sufficient data availability not only on outcomes but also (importantly) on implementation activities, the presence of stakeholders committed to an evaluation, and (importantly) whether the benefits of the evaluation justify the investment of resources. Of course, all of those pieces may not be present and an IE could still be appropriate, depending on whether the intention is to conduct a formative or summative evaluation, along with a more precise goal (i.e., Innovating, Pilot-Scale, Complementary, Confirmatory, Generalizing, or Optimizing). Critically, if USAID moves ahead with an IE with one of these aims, it should commit to all the necessary parameters to achieve a high-quality evaluation. If it cannot, then it may want to pursue another type of evaluation, such as a PE. The evaluability assessment should be revisited during the design process as more information becomes available.

CLARIFY STAKEHOLDER ROLES AT THE CONTRACT STAGE AND INCLUDE SPECIFIC PROVISIONS IN ALL CONTRACTING

To address the challenges associated with different stakeholder constraints and incentives, USAID should develop standard IE scope of work language to be used as a template in IP bidding documents for the USAID program and in Evaluator/PI bidding documents for the IE. Once a specific decision is made to conduct an IE, that language should be adapted for the needs of a particular intervention/evaluation and tailored to a stakeholder. It is possible to include sufficient detail without completing the IE design, as the contracting language is designed to set expectations and provide structure to the stakeholder interaction that follows. If a forced marriage is the analogy, then better contracting is akin to a prenuptial agreement designed to provide better structure, appropriate constraints, and optimal incentives.

[33] This is also a recommendation of Hageboeck et al. Impact Evaluation: Critical Challenges/Promising Solutions and consistent with PPL guidance.

At minimum, official contracting for IPs should include: 1) details on the expected intervention and evaluation designs; 2) leadership level of effort expectations; 3) project budget expectations, including the share of the budget connected to the evaluation; 4) information on expected assignment of program activities to beneficiaries, including the need to assign consistent with the IE design; 5) expectations on implementation fidelity, including commitment to implementation stability where required or implementation adaptation as needed; 6) monitoring and reporting requirements, especially detailed information about implementation, which evaluators need to characterize the administration of treatment; 7) coordination/communication mechanisms for regular engagement with evaluators; 8) staffing requirements, including an IE point of contact within the IP; and 9) mandatory post-evaluation action plans.

For evaluators, including connected academic PIs, parallel language needs to be developed, as well as some additional components. Initial contracts need to include provisions for 1) expectations of evaluator presence continually over the life of the project and not only during the baseline, midline, and endline; 2) evaluation contractors to serve as technical experts and not only contracting mechanisms; 3) evaluator disclosures to academic PIs; 4) coordination/communication mechanisms with USAID and IPs; 5) strategies to incorporate IPs into evaluation design and implementation decision-making; 6) strategies to incorporate local knowledge into evaluation designs and ensure adequate field presence; and 7) mandatory dissemination reports not only after endline, but about lessons learned over the lifecycle of the evaluation.

CONDUCT IEs AS TEAMS THAT INCLUDE THE IP

To address various lessons learned about communication and coordination, USAID should shift its thinking away from strict independence of programming and evaluation toward a collaborative approach in which the IP is included on the evaluation team, and an evaluation team member is engaged in a more sustained way throughout the program. Indeed, the World Bank employs a variation of this model, whereby evaluators (academic PIs and evaluator organizations) and the government (the implementer) are increasingly working closely over the life of the program and evaluation. This is in contrast to the conventional wisdom on IEs, referred to by one interviewee as the "conduct baseline and see you in 5 years" approach and another as the goal of a "firewall" between evaluators and programmers. A similar proposal was made by Hageboeck et al., who recommend a robust post-IP contracting workshop and IP participation in scoping and design, and Mohan et al. also recommend involvement of implementer staff and internal MEL experts in design and implementation.[34]

As just discussed above, a key step is to put language in requests for proposals (RFPs), and then solidify in contracting, that the IP should have as key personnel someone that can serve as a point of contact for the evaluation and engage on the evaluation team. This person would have other MEL or CLA responsibilities for the IP but should have adequate qualifications to understand the basics of an IE design and participate effectively in the evaluation design conversations, scoping trip activities, monitoring implementation fidelity, and coordinating with evaluators. In similar fashion, the evaluation team also needs a staff member in-country over the entire life of the project to coordinate with the IP, monitor implementation fidelity, oversee data collection, and liaise with stakeholders on a part time basis.

More broadly, the prior DRG Center IE Clinic model had many virtues, and the team recommends its continuation; however, it needs some critical adaptations, including getting IPs into the discussion much

[34] Hageboeck et al. Impact Evaluation: Critical Challenges/Promising Solutions; Mohan et al. Analysis of the Systematic Implementation of Rigorous Impact Evaluations.

earlier in the process, instead of just Missions and PIs. The past IE Clinic model roughly followed a process whereby scoping and design occurred prior to implementer onboarding. Instead, USAID should alter the process so that a clinic is geared toward considering design options to adequately inform an RFP, but scoping and the full design should occur after IP onboarding. Once the IP is onboarded, the evaluation team, which would then include a member from the IP, would engage in a multi-day IE design workshop, followed by scoping, and co-creation of a final design and MOU, whereby stakeholder responsibilities would be outlined and agreed to. This revised approach has some advantages, especially in engaging the IP in the evaluation design, but it does lead to some timing challenges in that the co-creation and finalization of design may occur while the implementer is beginning to develop programming. USAID will need to provide clear contracting language to the IP so that the IP does not proceed with programming prior to the completion of the evaluation design and the collection of baseline data.

And yet, even an improved IE Clinic model does not ensure that communication and coordination will continue throughout a program. Just as importantly, or perhaps more so, stakeholders need a mechanism for meaningful engagement on a regular basis. As discussed earlier, this includes regular meetings, shared folders, and formal or informal agreements on information sharing. In the earliest stages, the engagement might be better oriented toward developing shared understandings of different approaches, incentives, and modalities. Later, the engagement could be more focused on working through the ongoing data challenges, or learning agendas, associated with implementation and evaluation.

This recommendation will require accommodation by all stakeholders, especially evaluators and PIs. First, evaluation teams need to see the evaluation as something much more dynamic, in which adjustments need to be made both to programming and evaluation. Second, evaluators need to update their approach to learning. The textbook approach of conducting a grand evaluation, disseminating the final results, and motivating later use is short-sighted. Instead, regular coordination allows evaluators to share lessons learned at any given point in the process, something helpful to implementers and the Mission making real-time decisions. And, for their part, implementers can share the complexities of program rollout so that the evaluation correctly tracks the intervention; this would be helpful for evaluators. As noted earlier, novel approaches pairing individuals from evaluator organizations with academic PIs, or senior and junior academic PIs, or the IE counterparts within the evaluator and implementer, could help facilitate a more dynamic approach to IEs, and make evaluation contractors as well as IPs a more central part of the IE.

USAID MISSIONS AND THE DRG CENTER SHOULD TAKE A STRONGER ROLE IN HARMONIZING STAKEHOLDERS

As discussed above in the Question 3 response, USAID—and the DRG Center in particular—has an important role to play in facilitating coordination among diverse stakeholders, which was at times undermined by inadequate staffing and limited bandwidth. In most cases, the IP is contracted by the Mission and the evaluators are contracted by the DRG Center, which means that the Mission and DRG staff need to coordinate effectively to reduce friction. Each stakeholder comes with distinct strengths and weaknesses, and each has its own incentives and constraints. Contracting can set the structure, and the Mission has a basic role to harmonize the various actors and encourage communication and coordination. Given DRG Center staff's methodological expertise and experience across multiple evaluations, the DRG Center should play a key role in reducing dysfunction that arises from inter-stakeholder conflict. Clarity on the evaluation purpose is again critical here, and it shapes the structure of the relationship and the flow of communication and coordination. Much of the time, the PI and evaluator need to provide direction for

the evaluation. At times, the IP may be best positioned. In either case, the Mission and DRG Center need to play active roles in helping construct, and then continually promote, harmonious relations.

UNDERTAKE A GREATER NUMBER OF IEs ON MORE INTERVENTIONS, BUT FOCUS THE SCOPE AND SHORT-CIRCUIT MORE OFTEN

Consistent with the need to conduct IEs that fulfill a variety of formative and summative purposes, USAID should encourage a more nimble and far-reaching IE approach. Large, multi-year, single-shot IEs are at times necessary for summative evaluations, but they are likely misused, or overused. Missions or other stakeholders may be better positioned to conduct strategically (or opportunistically) formative evaluations when needed or where possible. Indeed, smaller and more targeted IEs that fulfill other functions (such as innovation, piloting, and complementary learning) could be usefully built into the component parts of many more programs. Efforts to create systematic adaptive experimental designs[35] may hold promise for *Innovating* and *Pilot IEs*, and be especially promising for organizations such as USAID that operate based on principles of adaptive programming. The core principle behind adaptive experimental design is to systematically evaluate a number of possible interventions to address a DRG problem and then focus on those that appear most promising. Certainly, summative IEs are still important and DRG may need to have a stronger part in ensuring that a broader learning agenda shapes those evaluations and is in turn shaped by them. But, given the broader lack of understanding of what works in the DRG space, undertaking a greater number of targeted formative evaluations could be key. In some cases, IPs could carry out small-scale evaluations of impact, and ideally all stakeholders would have an impact-learning mindset. A dedicated evaluation team should lead most assessments of impact, however, with the evaluators including IP input on the team, but at the same time preserving independence.

MAKE DISSEMINATION AND USE CONSISTENT WITH IE PURPOSE

Dissemination and use strategies should be tailored to the purpose of the IE and their success should be measured accordingly.

INCORPORATE DISSEMINATION AND UTILIZATION THROUGHOUT THE IE LIFE CYCLE

A key step toward better dissemination and utilization is to incorporate both throughout the life of the IE, not simply at the end; this is likely more feasible in formative IEs, but also possible in summative IEs. Utilization should begin at the design stage. An early best practice of DRG IEs was to commission an evidence review of the existing literature and knowledge gaps. This evidence review at the onset of an evaluation could be used to influence the program at the earliest program design and RFP phase, and could be easily shared with other Missions who are designing programs in similar areas to ensure their designs are rooted in the existing evidence. Coupling evidence reviews with an evaluability assessment would help set objectives and guide programmatic and evaluation decisions.

The baseline survey presents another opportunity for use. IPs and USAID have an opportunity to collaborate with the IP team to suggest survey questions that can provide valuable context or logistical details, such as GPS data points, that can inform program implementation. Data can be shared with the IP

[35] Offer-Westort, Molly, Alex Coppock, and Donald P. Green. 2021. "Adaptive Experimental Design: Prospects and Applications in Political Science." American Journal of Political Science.

MEL team and used to inform baseline values and set targets in the program's MEL plan. Midline and endline data can also be used to supplement project MEL data.

After baseline data—and midline data, if it exists—is analyzed, there is another opportunity for the stakeholder groups to come together and re-visit IE learning questions. Consistent with USAID's Collaborating, Learning, and Adapting approach, if IEs are able to pivot once it is determined that difficulties in program design or implementation make it unlikely for the IE to find meaningful, useable results, the core stakeholders have an opportunity to adapt and develop new learning questions that can be answered. This was done successfully in Uganda, but was not attempted in Tanzania or South Africa, leading to IEs with very little utilization. An adaptive approach will require the evaluation team to be open to course changes in programming and will require USAID and the IP to be open to mid-program course correction that may change project budgets or timelines.

Finally, one approach the World Bank's IE team employs to increase utilization of the rich datasets IEs produce is to invite country governments and program implementers to send the IE team questions that are not directly related to the IE but are of interest to them and can be answered by working with the data. USAID could adopt a similar practice after each dataset is finalized to get more use out of an already-existing IE output and increase the relevance of the IE to the IP and other program stakeholders. These invitations for questions could take place at any time in or after the project life cycle and could help maintain an interest in and an appetite for the IE results. This would require building in some flexibility in IE budgets to allow for such adaptation and allocate time to activities that will increase evaluation utilization.

INCREASE THE ACCESSIBILITY AND ACTIONABILITY OF THE FINDINGS REPORT

The ubiquity of the findings report makes it a logical starting place to increase the utilization of IE findings. To make the document more accessible to a non-technical audience, USAID should ensure that each technical findings report is accompanied by a policy-oriented brief, complete with graphics and other data visualizations, that will help translate the findings to a less technical audience. To increase accessibility to the host country audience, the executive summary of the report and evaluation briefs should be translated into local languages.

The second recommendation to improve the findings report is to increase the actionable recommendations in the report. Increased communication and collaboration between the stakeholders will make it easier for the evaluation team to craft recommendations that are relevant to USAID and the IPs. For example, in the Uganda case, USAID/Uganda brought the evaluation team back to Uganda to speak with the IP and program beneficiaries with the specific goal of creating recommendations that could immediately influence program implementation. This resulted in guidance on what types of recruitment strategies would be most beneficial for reaching underrepresented target audiences, among others.

Effort should be devoted to using novel dissemination strategies. Rather than just standard approaches, such as the two-pager, new and effective means for dissemination need to be identified. Some World Bank units have used models such as the "Radically Brief, Policy Brief" or point-counterpoint "Smackdowns."[36]

[36] See, for examples, "Do the Poor Waste Cash Transfers: Evidence from 11 Countries Suggests No." Available online at: http://documents1.worldbank.org/curated/en/793251468188679810/pdf/98627-BRI-PUBLIC-ADD-SERIES-Box393179B-

Other approaches common in the private sector, such as explainer videos, are not used much in the development space but appear to be highly effective and are becoming much less expensive.

INVOLVE USAID STAFF IN CRAFTING RECOMMENDATIONS FOR USAID STRATEGY AND PROGRAMMING

There is a limit to what PIs can be expected to understand about USAID programming more widely, or USAID country or sector strategy. USAID staff, likely the program and/or evaluation contracting officer representatives (COR), are better placed to create recommendations that speak to broader USAID strategy and use. The MCC recognizes that internal staff are often best placed to make recommendations about how to use IE findings more broadly across the agency and, as a standard practice, requires each IE COR to draft a "Lessons Learned" document that identifies programmatic and evaluation lessons building on each evaluation. A similar process at the DRG Center would help ensure IE recommendations are applicable more broadly across countries and programs and would increase the IE's value to USAID.

Beyond dissemination strategies, there should also be a use strategy that includes potential uses during the course of the evaluation and after. As required by the ADS, post-evaluation action plans should be developed for all IEs. As above, the use will vary across IE purposes. In some cases, the primary users might be the IP or the Mission, while in other cases, the users might be far broader. In the former case, the IP or the Mission should be responsible for developing the required post-evaluation action plan; in the latter case, the DRG Center itself might be responsible for such a plan.

CREATE A CENTRAL REPOSITORY FOR POSTING REPORTS, EVIDENCE REVIEWS, DATASETS, POLICY BRIEFS, AND OTHER IE MATERIALS TO INCREASE ACCESS TO IE FINDINGS

A common refrain in interviews and in the survey was that stakeholders simply did not know what information existed or how it was utilized. USAID Missions were unaware of other DRG Center IEs on similar topics, such as increasing revenue collection or increasing accountability. Reports, briefs, and datasets on the DEC and DDL are difficult for the general public and USAID staff alike to access and utilize, missing an opportunity for wider learning from IE results. Creating a single online repository for all DRG IEs—and potentially PEs and other learning—that is easily searchable by IE, sector, or region would markedly increase the audience and shelf life for IE findings.[37] Furthermore, since interaction with the site could be measured, it would provide a metric for the size and scope of engagement with the materials.

While the DEC currently hosts the IE reports, the DEC is aptly named as a clearinghouse rather than a means to curate knowledge and learning. Being able to take a broad look at the evidence would allow the DRG Center to understand when common interventions need summative IEs, such as Confirmatory, Generalizing, or Optimizing, that require multi-country coordination and would otherwise be difficult without the Center. The DRG Center would also be in a position to identify and encourage innovative IEs to assess new ideas not on the radar of individual Missions that may be focused on a more narrow set

Aug2015.pdf and "Smackdown: Provide the People of Africa with Training, or with Cold Hard Cash?" Available online at: https://blogs.worldbank.org/africacan/smackdown-provide-the-people-of-africa-with-training-or-with-cold-hard-cash.

[37] USAID's Land and Urban Office's www.land-links.org is an excellent example of this type of platform. Other examples include Agrilinks and Edulinks.

of programming approaches. And Missions, for their part, could draw on the repository to inform their unique evaluation needs.

A centralized hub could also extend the lessons learned well beyond the effectiveness of interventions on outcomes through IEs. It could gather best practices for the actual conduct of the evaluations; for example, learning how to conduct Pilot-Scale IEs better such that the pilots occur in a sufficiently timely way to inform scale *within* a program cycle. Some centralization of lessons learned may be especially important as pillar or regional bureaus would be more likely to put IE findings to use in ongoing or future work; however, absent dissemination and coordination, they would not be aware.

If the DRG Center moves ahead with a central repository, it needs to commit credibly to the difficult tasks of 1) gathering decentralized information, 2) organizing the information according to a schema (that would need to be developed), 3) posting new information regularly to keep relevant, 4) commissioning regular reviews of the evolving evidence, 5) disseminating lessons learned from evaluating the evaluations, and, most importantly, 6) tirelessly pushing for use of the evidence.

Finally, the DRG Center should not seek to organize evaluation learning across other donors, but a site housed within the DRG Center would greatly facilitate the efforts of other organizations, such as 3ie (the International Initiative for Impact Evaluation), to gather evidence across donors for the DRG sector.

INCLUDE RESEARCH PARTICIPANTS AND LOCAL COMMUNITIES IN DISSEMINATION EFFORTS

Absent from dissemination audiences are the subjects of the research themselves. There is considerable opportunity for USAID to invest in ways to share research results with communities and individuals who made the research possible. Research that is committed to community participation can also mitigate the problematic potential of research to be "extractive" from the standpoint of the study population, as appropriate data dissemination can reduce the gap between researchers and communities by building trust and including communities in research benefits.[38] Such activities also have the potential to augment research projects by strengthening the rigor, relevance, and reach of such research.[39] Interaction between the community and researchers allows community members to use findings to produce programming adjustments and increase confidence in unanticipated results. It also opens up the possibility of communities using the research findings to make decisions or change their behavior in a meaningful way.

Also absent with a few exceptions (e.g., Ghana) are local universities and academics. Potential opportunities for sharing IE findings include presenting results to university political science and economics departments and sharing the data, once publicly available, with graduate students and professors to encourage them to use the data for their own research. Including academics from local universities on the PI team will further help build an audience and appetite for greater IE utilization. Although it is more common to include local IPs or survey firms, most DRG Center programs only include academic PIs and evaluators from the United States. Dissemination and inclusion of local academic partners could build medium- and long-term capacity for in-country evaluation contributions. Such engagement would be

[38] McDavitt, Bryce, Laura M. Bogart, Matt G. Mutchler, Glenn J. Wagner, Harold D. Green, Sean Jamar Lawrence, Kieta D. Mutepfa, and Kelsey A. Nogg. "Dissemination as Dialogue: Building Trust and Sharing Research Findings Through Community Engagement." Preventing Chronic Disease 13 (March 17, 2016). https://doi.org/10.5888/pcd13.150473.

[39] Balazs, Carolina L., and Rachel Morello-Frosch. "The Three Rs: How Community-Based Participatory Research Strengthens the Rigor, Relevance, and Reach of Science." Environmental Justice 6, no. 1 (February 2013): 9–16. doi:10.1089/env.2012.0017.

helpful for evaluation reasons and may also spur new lines of research for which local researchers could provide intellectual leadership.

7. CONCLUSION

Over a decade since the National Academy of Sciences outlined key roles for IEs in democracy assistance, the DRG Center has carried out 27 IEs and this retrospective takes stock of the challenges and lessons learned. Although USAID has a distinct approach, with accompanying strengths and challenges, its experience with IEs is not altogether unique. Indeed, a variety of other USAID bureaus, US Government agencies, and foreign donors, have all used IEs and are reflecting on its role in international development. Given the resources invested in IEs, and perhaps more importantly the resources invested in particular development programs that IEs are designed to assess, it is imperative to take a critical look.

Based on scores of interviews, a survey, and a desk review of all DRG IEs conducted to date, the retrospective outlines the key findings of IEs, challenges and lessons learned, dissemination and use, and provides a set of recommendations. Taken together, the report concludes that to realize the full potential of IEs, the prior DRG model needs substantial updating. Importantly, the necessary revisions are feasible and consistent with USAID's broader approach to IEs.

The DRG Center is at an important juncture with respect to IEs, and the report's findings indicate that IEs should be a key part of DRG Center activities. There is much to learn about the effects of DRG activities, and many of those activities are, in fact, evaluable. With attention to the objectives and design of IEs, the DRG Center is well-positioned to advance a learning agenda that promises to improve its own democracy, human rights, and governance programming, and that of the broader development community.

ANNEX 1. THREE GENERATIONS OF DRG CENTER IEs (N=29)

GENERATION	DESCRIPTION	REGION	SECTOR	COMPLETED	COST/ BUDGET	CONTRACT	TASKING	SURVEY RESPONSES
3	**Bangladesh — Countering violent extremism**	Asia	Justice and Rights	Completed	$485,513	DRG-LER I	N058	7
3	Burkina Faso— Countering violent extremism	Africa	Justice and Rights	Completed	$600,560	DRG-LER I	N062	5
1	**Cambodia - Constituency dialogues**	Asia	Democratic and Political Processes	Completed	.	CEPPS- NDI	.	1
1	Cambodia - Youth civic engagement	Asia	Civic Power and Citizen Engagement	Completed	.	CEPPS- IRI	.	1
3	Cambodia- C-TIP	Asia	Justice and Rights	Completed	$684,972	DRG-LER I	N040	2
3	Cambodia— Women's political participation	Asia	Civic Power and Citizen Engagement	Completed	$696,153	DRG-LER I	N020	
1	Central America (Not on DEC)	Latin America Car	Justice and Rights	Completed	.	.	.	
1	Dominican Republic- Program for Political Education and Management (Not on DEC)	Latin America Car	Civic Power and Citizen Engagement	Completed	.	.	.	
3	DRC— Gender-based violence	Africa	Justice and Rights	Completed	$1,048,978	DRG-LER I	N059	1
3	DRC— Integrated governance	Africa	Governance and Financing Self Reliance	Completed	$1,518,478	DRG-LER I	N035	2
3	**Eastern and Southern Caribbean - Youth violence prevention**	Latin America Car	Justice and Rights	Completed	$2,533,589	DRG-LER I	N049	3

GENERATION	DESCRIPTION	REGION	SECTOR	COMPLETED	COST/ BUDGET	CONTRACT	TASKING	SURVEY RESPONSES
3	Georgia—Civic education	Eastern Europe	Civic Power and Citizen Engagement	Completed	$557,582	DRG-LER I	N015	6
2	**Ghana—Local government accountability**	Africa	Governance and Financing Self Reliance	Completed	$777,582	EDGE/ DRG-LER I	S002	6
3	**Haiti—PROJUSTICE legal support to pretrial detainees**	Latin America Car	Justice and Rights	Completed	$420,765	DRG-LER I	N031	3
3	Liberia - Candidate debates	Africa	Democratic and Political Processes	Completed	.	DRG-LER I	N072	3
3	**Malawi—Local resource mobilization**	Africa	Governance and Financing Self Reliance	Completed	$699,411	DRG-LER I	N030	5
3	Mali—Rule of law education	Africa	Justice and Rights	Completed	$393,835	DRG-LER I	N032	3
2	Mozambique — Voter participation	Africa	Democratic and Political Processes	Completed	$230,929	EDGE/ DRG-LER I	S005	
3	Nepal—Electoral participation	Asia	Democratic and Political Processes	Completed	$540,000	DRG-LER I	N033	3
3	Niger - Participatory and responsive government	Africa	Governance and Financing Self Reliance	Completed	$267,000**	DRG-LER I	N043	2
3	Paraguay—Integrated value chains	Latin America Car	Governance and Financing Self Reliance	Completed	$442,536	DRG-LER I	N018	4
3	Peru - Anti-corruption	Latin America Car	Governance and Financing Self Reliance	Completed	$317,296*	DRG-LER I	N016	2
1	Russia - Golos election observation	Eastern Europe	Democratic and Political Processes	Completed	.	.	.	

GENERATION	DESCRIPTION	REGION	SECTOR	COMPLETED	COST/ BUDGET	CONTRACT	TASKING	SURVEY RESPONSES
2	South Africa—Rape crisis center utilization	Africa	Justice and Rights	Completed	$501,856	EDGE/ DRG-LER I	S004	2
3	**Tanzania—Journalism training**	Africa	Civic Power and Citizen Engagement	In progress, close to complete	$600,560*	DRG-LER I	N061	3
3	Uganda - Radio experiment	Africa	Governance and Financing Self Reliance	Completed	$250,000**	DRG-LER I	N054	
2	**Uganda—SMS local government accountability**	Africa	Governance and Financing Self Reliance	Completed	$500,321	EDGE/ DRG-LER I	S008	4
3	Zambia—Parliamentary scorecard	Africa	Democratic and Political Processes	Completed	$317,219	DRG-LER I	N027	1
2	Zimbabwe—Supporting traditional leaders to mitigate conflict	Africa	Justice and Rights	Completed	.	EDGE		

Note: Case studies appear as emboldened text

* Budgeted amount

**Data collection only or otherwise excluded from cost calculations in the report.

ANNEX 2. CASE STUDY SELECTION

The case selection methodology went through several iterations. Several variables were identified as important selection factors with a final focus on the generation of the IE, the theory of change, the nature of the results achieved, the extent of implementation challenges, and utilization. LPs involved in NORC and Social Impact, IEs were asked to rate each evaluation on a three category ordinal scale to aid in case selection; however, the scoring process was highly subjective and where there were overlapping assessments there was low inter-coder reliability. Furthermore, there was a basic problem of too many variables and too few cases to allow for statistical control. In the end, there were cases from all three generations, only cases of moderate or strong theory of change were selected, two cases had strong utilization, one during the evaluation (Uganda) and one based on the findings (Haiti), several cases had implementation challenges (Malawi, ESC, Tanzania, Ghana), and cases varied in their results (null, mixed, positive).

DESCRIPTION	SELECTION CRITERIA AND CONSIDERATIONS
Ghana— Local government accountability	Strong theory of change, mixed results, moderate implementation challenges, some utilization Flagship 2nd generation IE
Uganda—SMS local government accountability	Moderate theory of change, null results, minor implementation challenges, strong utilization Example of utilization and influence during the course of the IE
Malawi—Local resource mobilization	Strong theory of change, mixed results, major implementation challenges, some utilization
Eastern and Southern Caribbean - Youth violence prevention	Strong theory of change, null results, major implementation challenges, utilization pending
Haiti— PROJUSTICE legal support to pretrial detainees	Strong theory of change, positive results, minor implementation challenges, strong utilization Example of strong utilization of findings
Bangladesh— Countering violent extremism	Strong theory of change, positive results, minor implementation challenge, utilization pending Example of subcontracted intervention
Tanzania— Journalism training	Moderate theory of change, null results, major implementation challenges, utilization pending
Cambodia— Constituency dialogues	Moderate theory of change, mixed results, moderate implementation challenges, some utilization. Example of first generation IE

ANNEX 3. LIST OF KIIS

	POSITION	COUNTRY/ REGION	IE
1	USAID/Operating unit	Bangladesh	Countering Violent Extremism
2	Implementing partner	Bangladesh	Countering Violent Extremism
3	Implementing partner	Bangladesh	Countering Violent Extremism
4	Implementing partner	Bangladesh	Countering Violent Extremism
5	Implementing partner	Bangladesh	Countering Violent Extremism
6	Implementing partner	Bangladesh	Countering Violent Extremism
7	Evaluator	Cambodia	Constituency Dialogues and Citizen Engagement in Cambodia: Findings from a mixed methods impact evaluation
8	Principal Investigator	Cambodia	Constituency Dialogues and Citizen Engagement in Cambodia: Findings from a mixed methods impact evaluation
9	Implementing partner	Cambodia	Constituency Dialogues and Citizen Engagement in Cambodia: Findings from a mixed methods impact evaluation
10	USAID/Operating unit	Caribbean	Youth Violence Prevention
11	USAID/Operating unit	Caribbean	Youth Violence Prevention
12	USAID/Operating unit	Haiti	Impact Evaluation of USAID Haiti PROJUSTICE Program Pretrial Detention Component
13	Principal Investigator	Malawi	Impact Evaluation of USAID/Malawi Local Government Accountability And Performance (LGAP) Activity
14	Evaluator	Tanzania	Impact Evaluation of USAID/Tanzania Media and Civil Society Strengthening Activity
15	USAID/Operating unit	Tanzania	Impact Evaluation of USAID/Tanzania Media and Civil Society Strengthening Activity
16	Principal Investigator	Bangladesh	Countering Violent Extremism
17	Principal Investigator	Bangladesh	Countering Violent Extremism

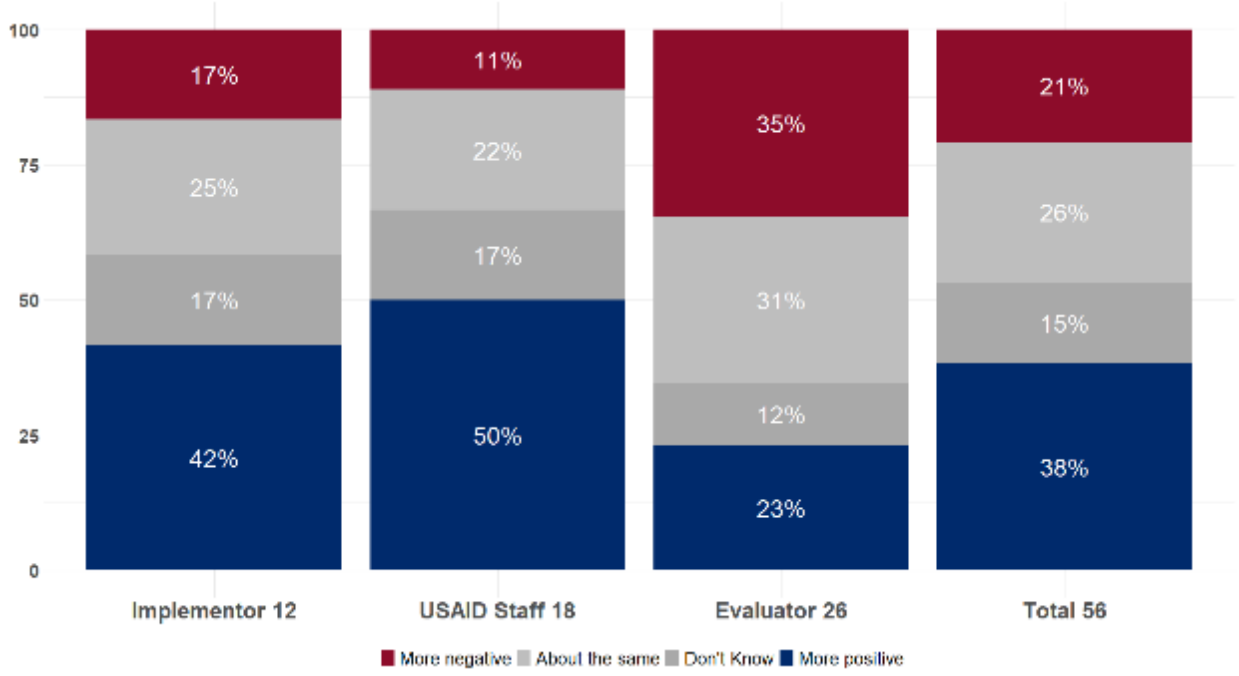
	POSITION	COUNTRY/ REGION	IE
18	Principal Investigator	Bangladesh	Countering Violent Extremism
19	USAID/Operating unit	Bangladesh	Countering Violent Extremism
20	Evaluator	Bangladesh	Countering Violent Extremism
21	USAID/Operating unit	Caribbean	Youth Violence Prevention
22	Evaluator	Caribbean	Youth Violence Prevention
23	Evaluator	Caribbean	Youth Violence Prevention
24	Principal Investigator	Caribbean	Youth Violence Prevention)
25	Principal Investigator	Caribbean	Youth Violence Prevention
26	Evaluator	Haiti	Impact Evaluation of USAID Haiti PROJUSTICE Program Pretrial Detention Component
27	Principal Investigator	Haiti	Impact Evaluation of USAID Haiti PROJUSTICE Program Pretrial Detention Component
38	Principal Investigator	Haiti	Impact Evaluation of USAID Haiti PROJUSTICE Program Pretrial Detention Component
39	USAID/Operating unit	Haiti	Impact Evaluation of USAID Haiti PROJUSTICE Program Pretrial Detention Component
30	USAID/Operating unit	Malawi	Impact Evaluation of USAID/Malawi Local Government Accountability And Performance (LGAP) Activity- Final Report
31	Evaluator	Malawi	Impact Evaluation of USAID/Malawi Local Government Accountability And Performance (LGAP) Activity- Final Report
32	Principal Investigator	Malawi	Impact Evaluation of USAID/Malawi Local Government Accountability And Performance (LGAP) Activity- Final Report
33	Principal Investigator	Malawi	Impact Evaluation of USAID/Malawi Local Government Accountability And Performance (LGAP) Activity- Final Report

	POSITION	COUNTRY/ REGION	IE
34	Implementing partner	Malawi	Impact Evaluation of USAID/Malawi Local Government Accountability And Performance (LGAP) Activity- Final Report
35	Evaluator	Tanzania	Impact Evaluation of USAID/Tanzania Media and Civil Society Strengthening Activity
36	Evaluator	Tanzania	Impact Evaluation of USAID/Tanzania Media and Civil Society Strengthening Activity
37	USAID/Operating unit	Tanzania	Impact Evaluation of USAID/Tanzania Media and Civil Society Strengthening Activity
38	Principal Investigator	Tanzania	Impact Evaluation of USAID/Tanzania Media and Civil Society Strengthening Activity
39	Principal Investigator	Tanzania	Impact Evaluation of USAID/Tanzania Media and Civil Society Strengthening Activity
40	Evaluator	Uganda	Endline Impact Evaluation Report of Governance, Accountability, Participation and Performance (GAPP): SMS for Better Service Provision in Uganda
41	Implementing partner	Uganda	Endline Impact Evaluation Report of Governance, Accountability, Participation and Performance (GAPP): SMS for Better Service Provision in Uganda
42	Principal Investigator	Uganda	Endline Impact Evaluation Report of Governance, Accountability, Participation and Performance (GAPP): SMS for Better Service Provision in Uganda
43	Principal Investigator	Uganda	Endline Impact Evaluation Report of Governance, Accountability, Participation and Performance (GAPP): SMS for Better Service Provision in Uganda
44	Principal Investigator	Uganda	Endline Impact Evaluation Report of Governance, Accountability, Participation and Performance (GAPP): SMS for Better Service Provision in Uganda
45	USAID/Operating unit	Uganda	Endline Impact Evaluation Report of Governance, Accountability, Participation and Performance (GAPP): SMS for Better Service Provision in Uganda

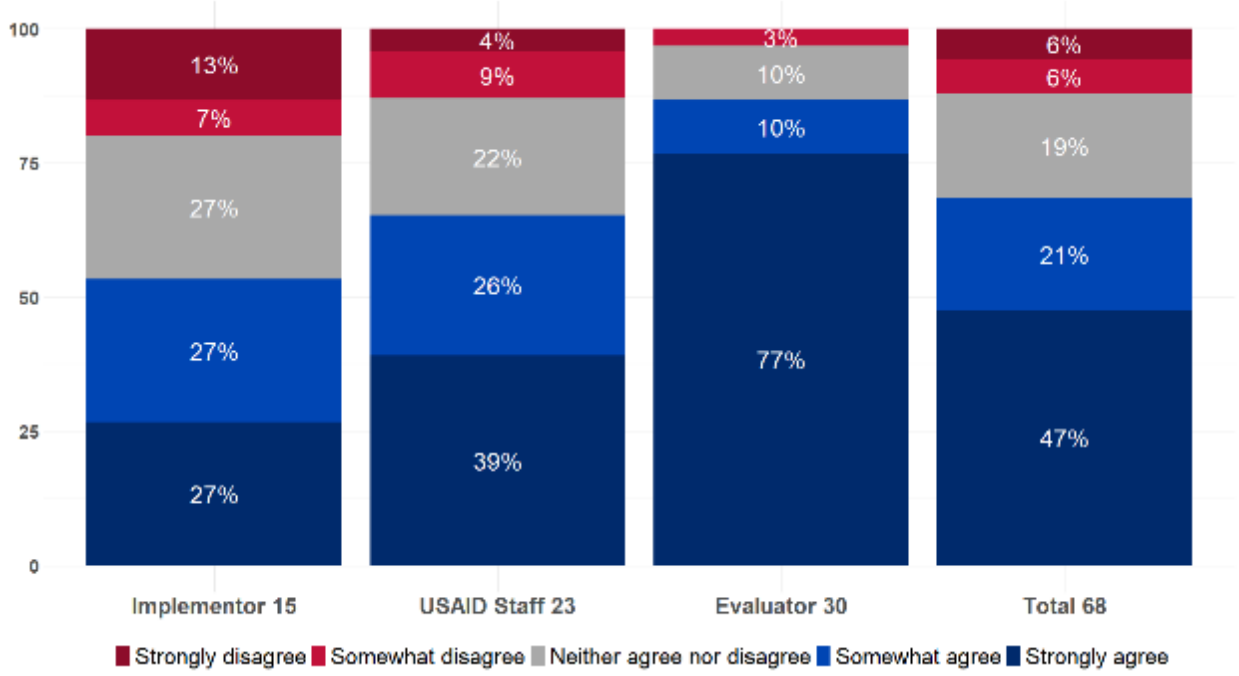
	POSITION	COUNTRY/ REGION	IE
46	USAID/DRG	Multiple	Multiple
47	Evaluator	Multiple	Multiple
48	USAID/Operating unit	Ghana	Endline Impact Evaluation Report: Ghana Strengthening Accountability Mechanisms (GSAM)
49	Principal Investigator	Ghana	Endline Impact Evaluation Report: Ghana Strengthening Accountability Mechanisms (GSAM)
50	Implementing partner	Ghana	Endline Impact Evaluation Report: Ghana Strengthening Accountability Mechanisms (GSAM)
51	Implementing partner	Ghana	Endline Impact Evaluation Report: Ghana Strengthening Accountability Mechanisms (GSAM)
52	USAID/Operating unit	Ghana	Endline Impact Evaluation Report: Ghana Strengthening Accountability Mechanisms (GSAM)
53	USAID DRG	Multiple	Multiple
54	USAID DRG	Multiple	Multiple
55	Evaluator	Multiple	Multiple
56	Evaluator	Multiple	Multiple
57	Evaluator	Multiple	Multiple
58	USAID/PPL	N/A	N/A
59	USAID/PPL	N/A	N/A
60	ILAB	N/A	N/A
61	ILAB	N/A	N/A
62	MCC	N/A	N/A
63	World Bank	N/A	N/A
64	World Bank	N/A	N/A

ANNEX 4. SURVEY RESULTS

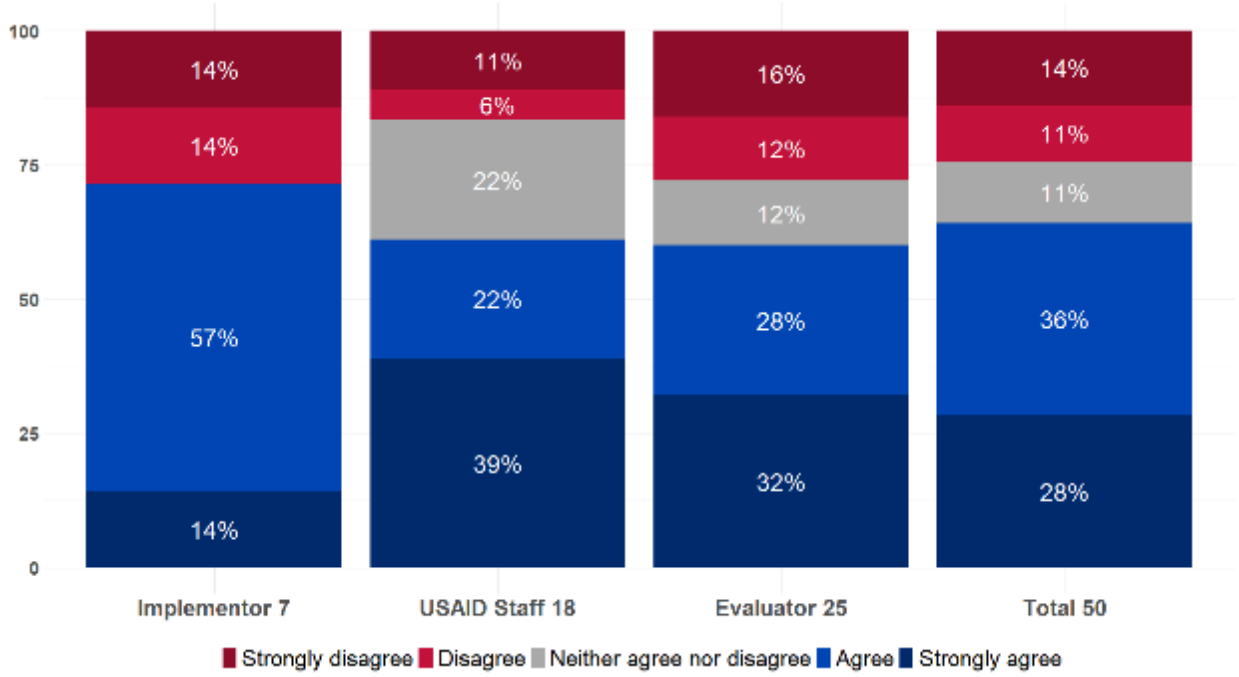
A6-Compared to pre-existing views on the impact of the project or activity at the time, how positive were the results?



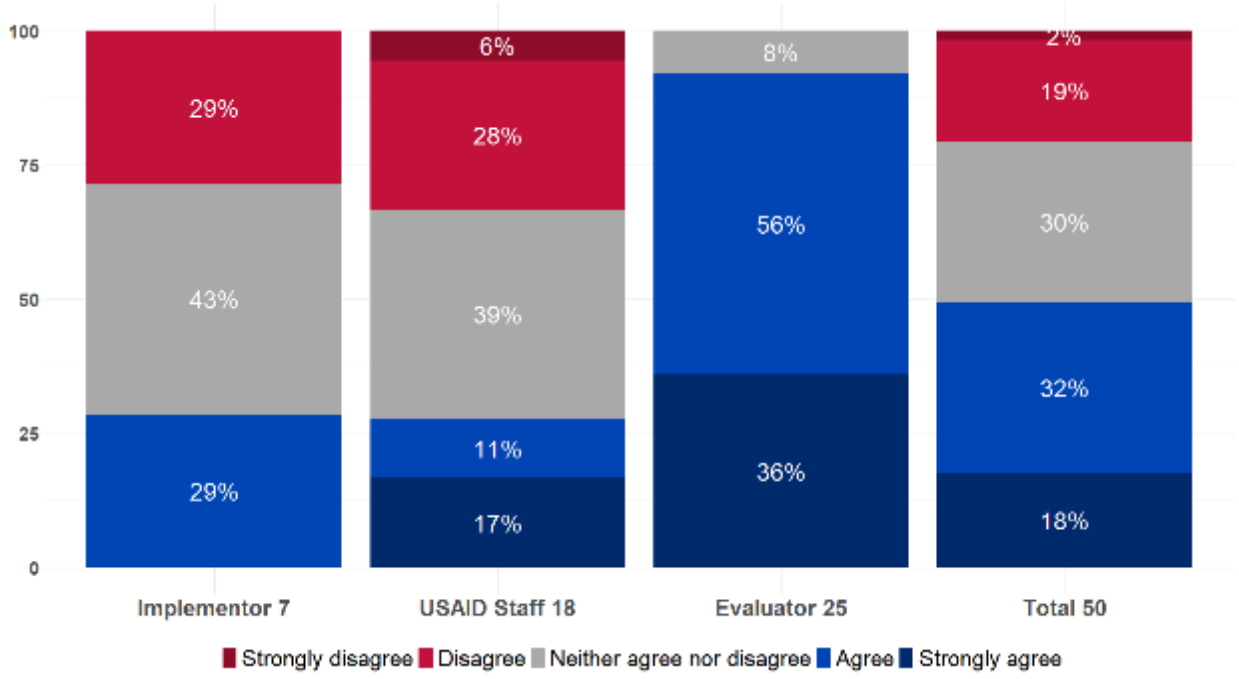
A7-Please rate how much you agree or disagree with the following statement: We learned more from the IE than could have been learned from more typical monitoring and a performance evaluation.



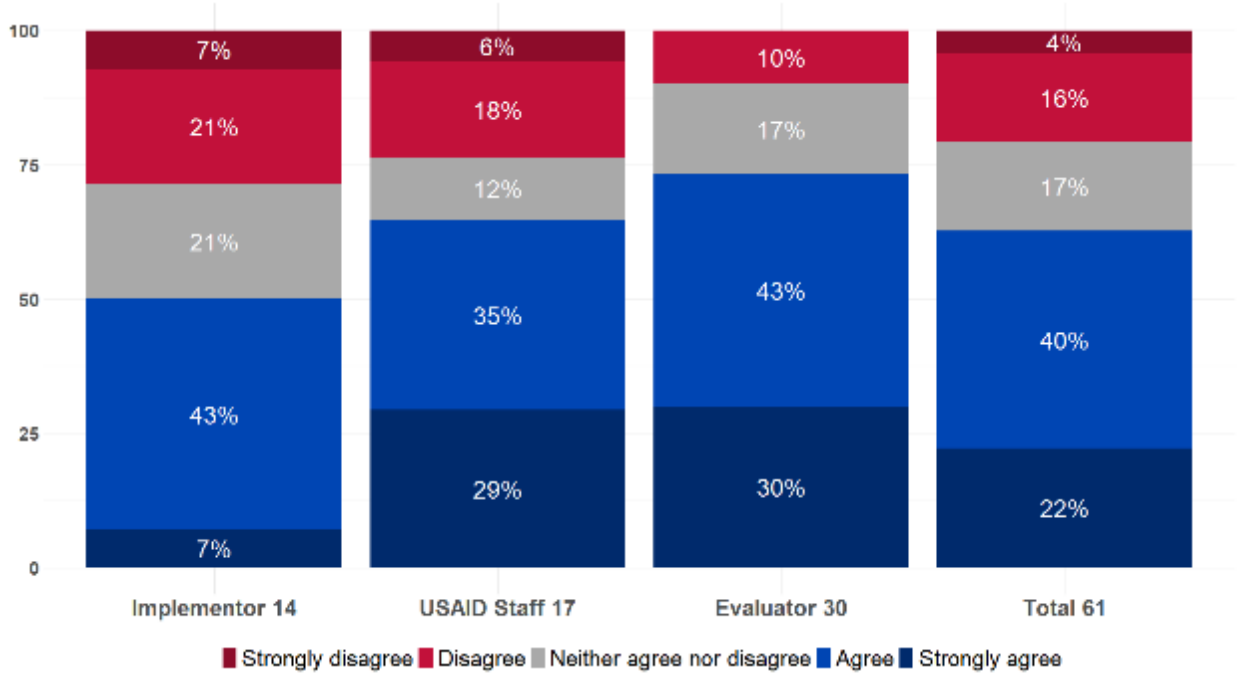
B3-Please rate how strongly you agree or disagree with the statement: The implementing partner was fully aware of the IE and the implications for program implementation at the procurement stage (RFP/RFA).



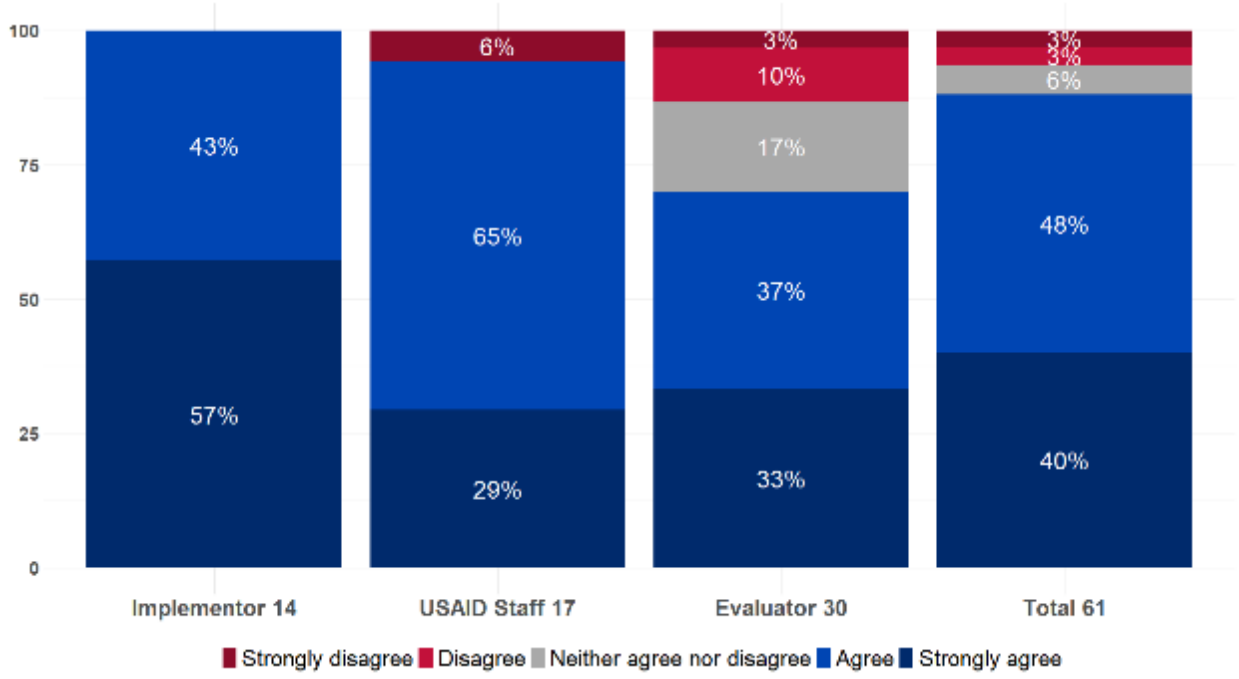
B4-Please rate how strongly you agree or disagree with the statement: The evaluation design matched the realities on the ground.



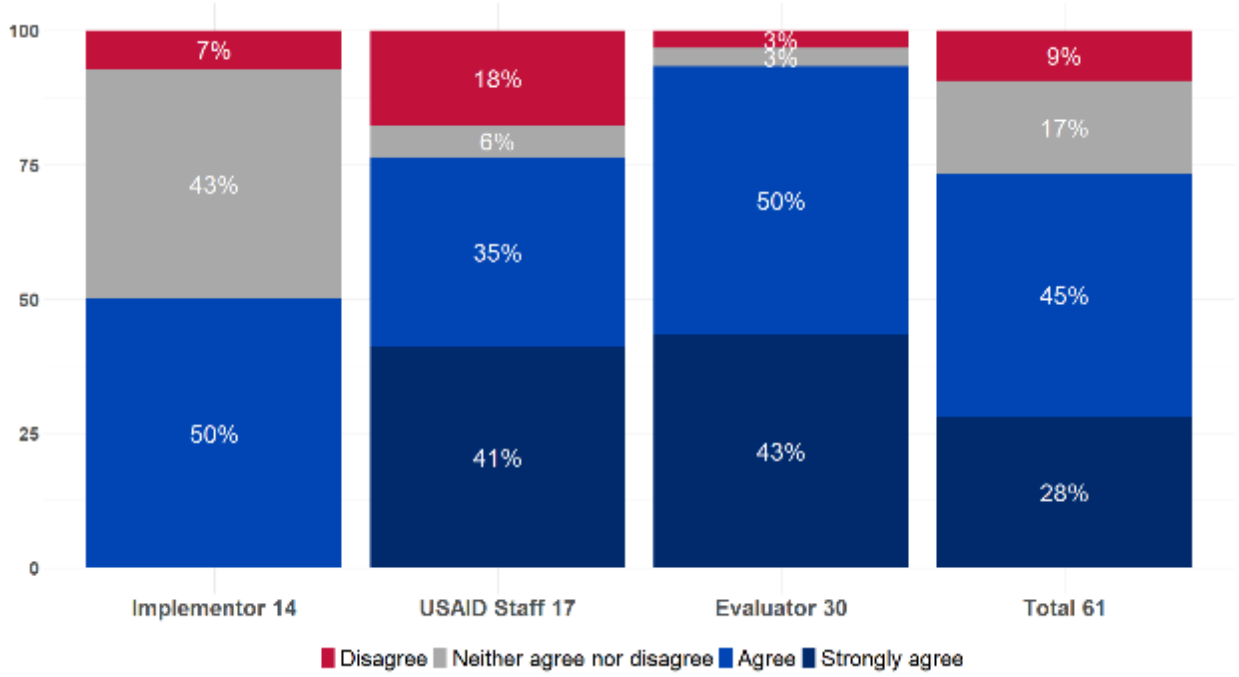
C2-Please rate how strongly you agree or disagree with the statement: USAID successfully balanced the needs of the evaluation and the needs of the intervention.



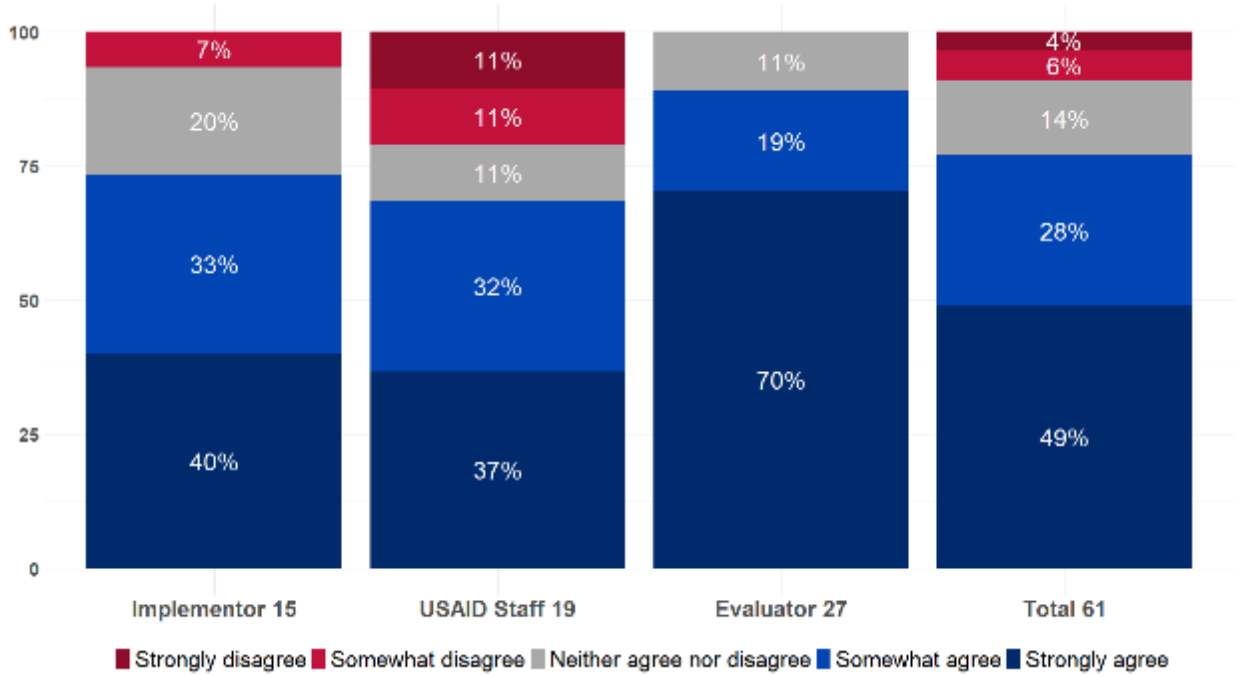
C3-Please rate how strongly you agree or disagree with the statement: The implementing partner was willing to accommodate the evaluation.



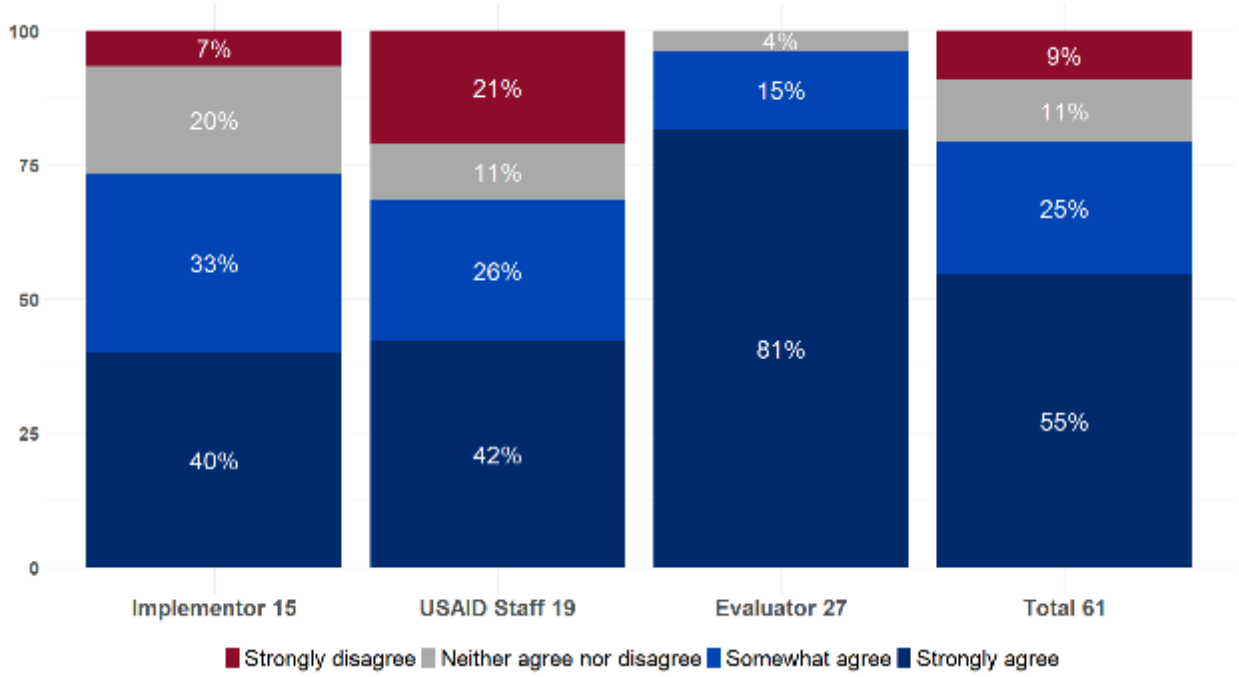
C4-Please rate how strongly you agree or disagree with the statement: The evaluation was sufficiently resourced to produce usable evidence.



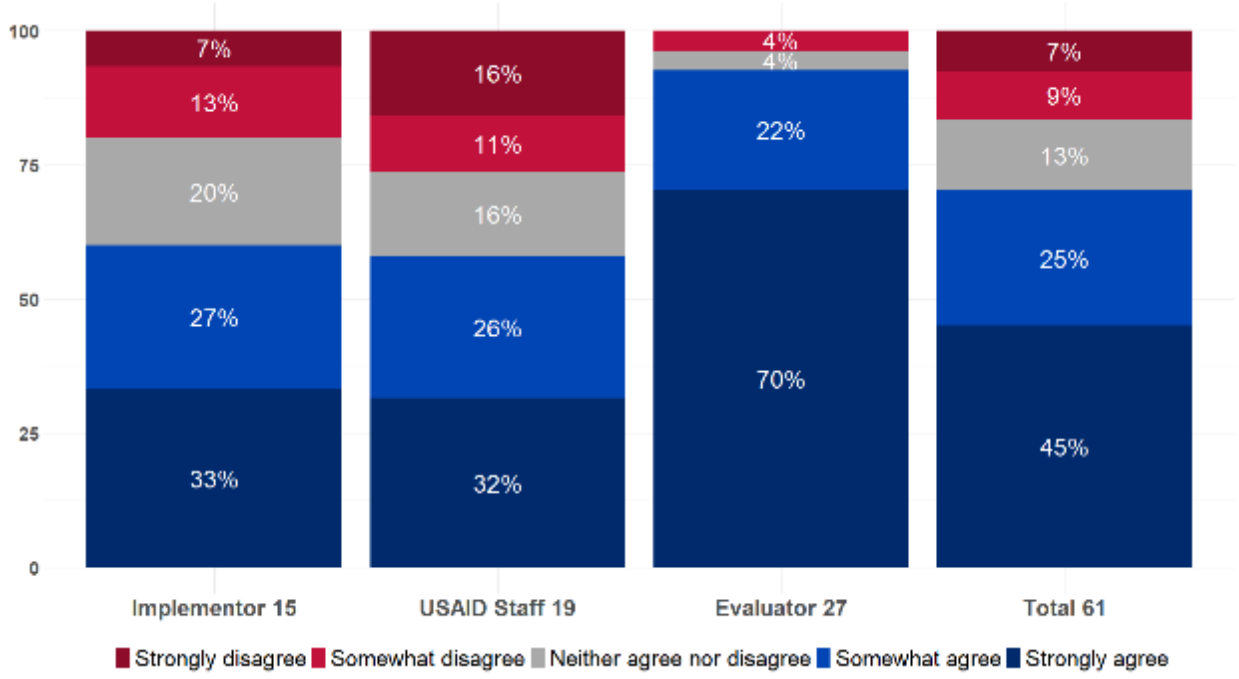
F2-To what extent would you agree with the following statements: I would encourage others at my organization to participate in an IE.



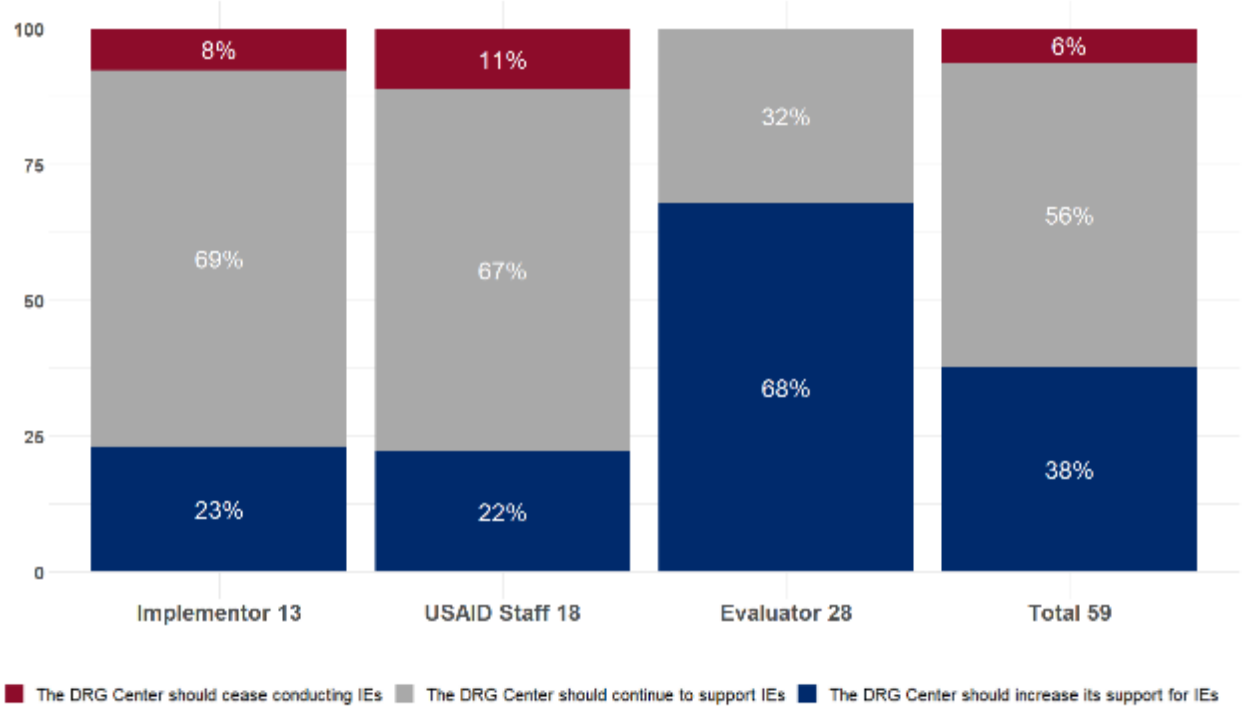
F3-To what extent would you agree with the following statements: I would participate in another IE if requested.



F4-To what extent would you agree with the following statements: I would actively seek out opportunities to participate in another IE.



F5-Which statement best represents your opinion about the future of USAID/DRG support for IEs?



ANNEX 5. COI DOCUMENTATION

Name	Daniel Sabet
Title	Senior Learning Advisor
Organization	USAID
Evaluation Position?	☐ Team Leader X **Team member**
Evaluation Award Number *(contract or other instrument)*	
USAID Project(s) Evaluated *(Include project name(s), implementer name(s) and award number(s), if applicable)*	DRG Center Impact Evaluation Retrospective
I have real or potential conflicts of interest to disclose.	**X Yes** ☐No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to:* 1. *Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated.* 2. *Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.* 3. *Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.* 4. *Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.* 5. *Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.* 6. *Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.*	I previously served as Chief of Party for learning partner Social Impact and was involved to varying degrees in six of the 27 IEs examined in this study. I currently serve as Senior Learning Advisor on the DRG Center's Evidence and Learning Team, which is responsible for the DRG Center's IE program.

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	
Date	3/11/21

ANNEX 6. RETROSPECTIVE SCOPE OF WORK

USAID DRG-LER II Tasking Request:
Tasking CB023: Impact Evaluation Retrospective

SUMMARY

This is a Statement of Work for a retrospective of impact evaluations (IEs) commissioned and conducted under the EDGE (2010-2014), DRG-LER (2013-present), and DRG-LER II (2018-present) mechanisms. This study will entail direct USAID participation on the retrospective team and a complete first draft is to be completed prior to a January 2021 learning event.

PURPOSE

Provide a retrospective of IEs conducted by the DRG Center that (1) provides an understanding of key challenges and lessons learned from both completed and uncompleted IEs, (2) assesses how the IEs have been used, and (3) provides recommendations for how to approach future evaluation to improve execution and utilization in the context of USAID DRG assistance.

INTENDED USE

The retrospective will be used by the DRG Center (1) as a summary document for reporting and reference purposes, (2) to inform decisions about future DRG evaluation work, (3) as a lessons learned document for future DRG Center evaluators and others conducting DRG impact evaluations.

BACKGROUND

In response to a 2008 National Academies of Science report, USAID's DRG Center initiated a pilot program of impact evaluations under EDGE (2010-2014), and additional DRG IEs were initiated under other mechanisms. The DRG Center's IE work was later systematized and scaled up under DRG-LER (2013-present) and DRG-LER II (2018-present). To date the DRG Center and its learning partners, in particular NORC and Social Impact, have completed 24 impact evaluations and have four ongoing. Impact evaluations have garnered substantial support among a core group of internal and external stakeholders; however, the DRG Center's IEs have also produced several critics frustrated with the challenging implementation process and concerned about IE usefulness. In 2019, the DRG Center began to scale back its impact evaluation work and initiated only two new potential impact evaluations in 2019 and 2020. This retrospective intends to provide a look back at both the accomplishments and the challenges of DRG Center impact evaluations with the goal of deriving lessons learned and recommendations for future evaluation work.

Note on the term retrospective: We use the term retrospective rather than evaluation because we are unable to commission an external evaluation of the Learning Team's own work. Nonetheless, frank and objective analysis and conclusions are expected and encouraged.

RETROSPECTIVE QUESTIONS

1. **Description**: How many impact evaluations were initiated; how many were completed; what was the cost of these evaluations, what topics and regions did they target? For those that were not completed, why were they not completed?
2. **Findings**: At a high level, what has USAID learned from the findings of these impact evaluations?[40]
3. **Challenges and lessons learned:** What have been the challenges encountered in designing and carrying out impact evaluations and what are the related lessons learned (for the DRG Center, Missions, implementing partners, and evaluators)?
4. **Use:** How has USAID (or others) used the impact evaluations? Why were some evaluations more useful than others?
5. **Recommendations**: What should be the DRG Center's approach to IEs moving forward? Under what conditions are they most effective and useful? How could the DRG Center better support Missions and others in the utilization of IE findings/recommendations?

RETROSPECTIVE METHODOLOGY

The Learning Partner is responsible for proposing a retrospective design. It is anticipated that the design will entail the following elements:

- Review of secondary literature focused on lessons learned in conducting impact evaluations in general and of DRG interventions in particular.
- Qualitative interviews with key current and former staff at the DRG Center, NORC, and Social Impact.
- Document review of the impact evaluations, including the Learning Harvest compilation of IE findings, 2-pagers, and evaluation reports. The methodology might involve coding of IEs for key variables, such as clarity of findings (e.g., negative, null, mixed, positive), implementer-related implementation problems (e.g., control group receives treatment, challenges adhering to timelines, funding issues), and evaluation-related implementation problems (e.g., randomization error, sampling error, measurement error, Mission or DRG Center staff turnover, funding issues).
- Either an online survey of the population of Principal Investigators (PIs), Mission points of contact (POCs), and implementer POCs, qualitative interviews with a sample of these individuals, or a mix of the two.

KEY PERSONNEL

Team Leader: The Team Leader must have extensive evaluation and impact evaluation experience. The Team Leader should be knowledgeable of USAID Evaluation Policy and have experience working with USAID and USAID Missions. Outstanding verbal and written English communication skills and qualitative research expertise are required. Experience with similar lessons learned or retrospective exercises is highly desirable.

DRG Center participant: To encourage retrospective use and to take advantage of institutional memory, the retrospective team will include a DRG Center participant. This is expected to be Daniel

40 The Learning Harvest provides summary findings for each completed evaluation. The answer to this question should be at a high level.

Sabet, current Senior Learning Advisor at the DRG Center and former Chief of Party for Social Impact under DRG-LER. The DRG Center participant will participate in the design and data collection, provide inputs to deliverables, and review drafts. Recognizing the potential for a conflict of interest (COI), the team leader will have final say over retrospective findings, conclusions, recommendations, and wording. Sabet's potential COI will be disclosed in an annex to the report.

Evaluation Specialist: The Specialist should have evaluation and impact evaluation experience. The Specialist should be knowledgeable of USAID Evaluation Policy and have experience working with USAID and USAID Missions. Strong oral and written English communication skills and qualitative research experience are required. Experience with similar lessons learned or retrospective exercises is desirable.

Deliverables: Deliverables are listed below and followed by a table with delivery dates. A strong draft of the retrospective must be completed prior to a late January 2021 Learning Event. As such the proposed methodology should be feasible to meet this deadline.

- ***Concept note***: The Concept Note should follow past Concept Note structure and include proposed staffing.
- ***Retrospective Workplan***: The Workplan should include among other elements (1) an explanation of the methodology, (2) an explanation of any sampling, (3) a list of interviewees or survey respondents, and (4) draft interview or survey instruments. Depending on the level of detail in the Concept Note, the Workplan could be a revision to the Concept Note.
- ***Draft Retrospective Report***: The report should be relatively brief at around 20 pages. More detailed analysis or information can be included in a report annex. An illustrative outline is included below.
- ***Final Retrospective Report***: In addition to a clean final report, the Learning Partner should submit a track-change version with responses to stakeholder feedback. Once approved the report should be 508 compliant and uploaded to the DEC.
- ***Two-pager or infographic***: The Learning Partner should develop a summary of the report's key findings and recommendations. This will be made available at the Learning event. Once approved this will be uploaded to the DEC along with the report.
- ***PowerPoint and Presentation***: The Learning Partner should also develop an accompanying PowerPoint that will be presented at the learning event.
- ***Updated report***: Of the four IEs in process, two should be completed in 2020 and the remaining two will be completed by September 2021. The Learning Partner should reserve some time to review these upon completion and update the report if necessary.

TABLE 1: TIMELINE

STEP OR DELIVERABLE	LEAD TIME	COMPLETION DATE
Issuance of Tasking		October 7, 2020
Concept Note	2 weeks	October 21, 2020
Approval	1 week	October 28, 2020
Retrospective Work Plan	2 weeks	November 11, 2020
Draft Report	7 weeks (+ 1 holiday week)	January 6, 2021
Comments on the Draft Report	2 weeks	January 20, 2021
Presentation at Learning Event; Two pager or infographic		TBD January, 2021
Final Retrospective Report	2 weeks	Feb 3, 2021
Updated Report		TBD September 2021

TABLE 2: ILLUSTRATIVE OUTLINE

SECTION	CONTENT
Executive summary (2 pages)	
Introduction (1-2 pages)	• Background including explanation of 2008 National Academy of Sciences recommendation, EDGE, DGR-LER, DRG-LER II. • Key elements of the DRG Center's approach to impact evaluation (e.g., testing theory, multiple treatment arms, design at the PAD stage, IE clinic)
Question 1: Description (2 pages)	• Number of IEs, regional distribution, costs • The number that were not able to move beyond the design stage and why.
Question 2: Findings (2 pages)	• High level highlights of what we have learned from the IEs (or not learned).
Question 3: Challenges and learning (4 pages)	• Challenges: Impact on intervention efficiency and quality, long-time horizon, randomization and design limitations, data and measurement limitations, etc.. • Lessons learned: Design at the PAD stage, need for robust theory of change, implementer-evaluator cooperation, key role of the PI, role of IE Clinics, etc…
Question 4: Use (3 pages)	• Existing programs • Future programming decisions • Changing understanding/culture/perspectives • Generating a common pool of evidence
Recommendations (2 pages)	
Annex	• Expanded Learning Harvest material for IEs with DEC links. For example, there could be a page for each IE with augmented tables from the Learning Harvest. • Instruments • COI documentation

ANNEX 7. QUANTITATIVE SURVEY INSTRUMENT

Thank you for your participation in this brief survey about your experience with impact evaluations conducted by USAID's DRG Center. The survey is a component of an IE Retrospective Evaluation commissioned by USAID/DRG that aims to inform decisions about future Democracy and Governance Center evaluation work and provide lessons learned for evaluators and others conducting DRG Impact Evaluations. In total, 135 contracting officer representatives, evaluators, and implementing partners from 27 different impact evaluations will be interviewed. All data and identifying information will be anonymized; it will be impossible to reconstruct your answers. Nobody will be identified by name, and it will be impossible to attribute any survey responses or findings to you.

This questionnaire will take about 20 minutes to complete. The final report will be made available in early 2021.

If you have any questions or concerns, please feel free to contact Dan Sabet at dsabet@usaid.gov.

SECTION A. IE INFORMATION		
A1.	Which evaluations were you involved in? *Select all that apply* *For each IE selected, ask 1b, c, and d*	
A2.	Which describes your role at the time of the impact evaluation?	1=USAID/DRG Center staff 3=USAID/Mission staff 3=Implementing partner/home office staff 4=Implementing partner/field staff 5=Evaluation partner staff (non-PI) 6=Principal investigator 97=Other, specify
A3	What is the status of the evaluation? *If 1 or 2, ask A3b*	1=Cancelled: Unable to design a viable evaluation 2=Cancelled: Evaluation was designed but the evaluation could not be completed 3=Ongoing 4=Completed
A3b.	Why was the evaluation cancelled?	*Text*
A4.	Which parts of the evaluation were you involved with? *Select all that apply*	1=Evaluation preparation and/or contracting 2=Evaluation design 3=Baseline data collection 4=Baseline analysis and reporting 5=Baseline dissemination or use 6=Midline data collection 7=Midline analysis and reporting 8=Endline data collection

		9=Endline analysis and reporting 10=Endline dissemination or use 11=Project implementation 12=Other, specify
A5	What program impacts did the IE find? *Ask only if A3=4*	1=Positive impact 2=Null (no) impact 3=Negative impact 4=Some positive impacts and some negative/null impacts 888=Don't know
A6	Compared to pre-existing views on the impact of the project or activity at the time, how positive were the evaluations results?	1=More positive 2=More negative 3=About the same 888=Don't know
A7.	What did you learn – or did you expect to learn – from this IE that you could not have learned from a traditional M&E or a performance evaluation?	*Text*
SECTION B. DESIGN AND PROCUREMENT *Ask if A4=1 or 2 or 97*		
B1.	Was this evaluation workshopped in the DRG Center's Impact Evaluation Clinic (a multi-day event with both Mission staff and academics to plan impact evaluations).	0=No 1=Yes 888=Don't know
Please rate how much you agree or disagree with the following statements		
B2.	The IE clinic helped ensure a good evaluation design. *If 1 or 2, ask B3. If 4 or 5, ask B4. If 3, skip to Be5*	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
B2b.	How did it help? *Skip to B5*	*Text*
B2c.	Why did it not help?	*Text*
B3.	The evaluation design informed the bidding documents for procuring the implementing partner	1=Strongly agree 2=Agree 3=Neither agree nor disagree

		4=Disagree 5=Strongly disagree
B4.	The evaluation design matched the realities on the ground *If 4 or 5, ask B5b.*	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
B5b.	How could the design have been improved to better match the realities on the ground?	Text
C. IMPLEMENTATION *Ask if A4=3, 4, 5, 6, 7, 8, 9, 10, 11, or 12*		
C1.	Impact evaluations often face challenges during their implementation. During which of the following stages did this IE experience challenges? *Select all that apply. For each challenge selected, ask C1b.*	1=Pre-procurement design (identifying an evaluable theory of change/learning question) 2=Design, including choosing methodology 3=Data collection 4=Data analysis and reporting 5=Dissemination/learning 97=Other, specify
C1b.	What was the challenge? What lessons were learned?	
Please rate how much you agree or disagree with the following statements		
C2.	USAID successfully balanced the needs of the evaluation and the needs of the intervention *If 4 or 5, ask C2b*	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
C2b.	How could this have been improved?	*text*
C3.	The implementing partner was willing to accommodate the evaluation *If 4 or 5, ask 32b*	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
C3b.	How could this have been improved?	*Text*
C4	The evaluation was sufficiently resourced to produce usable evidence	1=Strongly agree 2=Agree

		3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
C5.	In order to design and carry out the evaluation, did the intervention need to be altered after the procurement/onboarding of the implementing partner? *If 1, 2, or 3, ask C6b,c, d,* *If 4 or 888, skip to C6.*	1=Yes, in significant ways 2=Yes, in moderate ways 3=Yes, in minor ways 4=No 888=Don't know
C5b.	What changes were made?	*Text*
C5c.	*Evaluator/PI/Evaluator COR only* What impact did these changes have on the quality of the IE?	1=Very positive 2=Positive 3=Some positive, some negative 4=Negative 5=Very negative 6=No impact 888=Don't know
C5d.	Please describe the impact	*Text*
C5e.	*IPs/USAID COR only* What impact did these changes have on the quality of the program implementation?	1=Very positive 2=Positive 3=Some positive, some negative 4=Negative 5=Very negative 6=No impact 888=Don't know
C5f.	Please describe the impact	Text
C6.	*IPs/USAID only* Did you or your team use the baseline findings, including the report or dataset, to inform or adapt the program? *If 1, ask C6b.*	0=No 1=Yes 888=Don't know 777=No baseline
C6b.	How did you use them?	*Text*
C7.	*IPs/USAID only* Did you or your team use the midline findings, including the report or dataset, to inform or adapt the program? *If 1, ask C7b.*	0=No 1=Yes 888=Don't know 777=No midline
C7b.	How did you use them?	*Text*

C8.	*IPs/USAID only* Did you or your team use the endline findings, including the report or dataset, to inform or adapt the program? *If 1, ask C8b.*	0=No 1=Yes 888=Don't know 777=No endline
C8b.	How did you use them?	*Text*

D. CONJOINT EXPERIMENT		
Please read the description of the following impact evaluations carefully. Please indicate which of the two impact evaluations you would use for future programming?		
Dimension	IE 1	IE 2
Null vs. non-null results	The IE finds that the program had expected effects	The IE finds that the program did not have expected effects
Timeliness of evaluation reports	Results received before next iteration of project is determined	Results received after next iteration of project is determined
The evaluation was conducted by	U.S. contractor	Local contractor
Evaluation report that has	Concrete recommendations	Implied recommendations
Method	Randomized	Non-randomized
RCT vs. non-RCT vs. qualitative	RCT evaluation	Quasi-experimental eval
Evaluation cost	5%	20%

Section E. Learning *Asked only of complete IEs (A3=4)*		
Please rate how much you agree or disagree with the following statements		
E1.	The final evaluation report was easy to understand	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree

		5=Strongly disagree
E2.	The final evaluation report was timely	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
E3.	The final evaluation report was widely read by my team	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
E4.	The final evaluation report had actionable recommendations	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
E5.	The final evaluation findings were shared widely and with the appropriate audiences	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
E6.	The IP used IE findings to make decisions about the program being evaluated	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
E7.	The IP used IE findings to make decisions about other programs (current or future)	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
E8.	USAID used IE findings to inform decision making on the program being evaluated	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
E9.	USAID used IE findings to inform decision making on future iterations of the program being evaluated	1=Strongly agree 2=Agree 3=Neither agree nor disagree

		4=Disagree 5=Strongly disagree
E10.	USAID used IE findings to inform decisions about similar programs in other Missions (current or future)	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
E11.	Were there any other ways the findings of the IE were used to influence program adaptation or agency learning? *If 1, ask E11b.*	0=No 1=Yes 888=Don't know
E11b.	What were they?	*Text*
E12.	What type of recommendations did the report make? *Select all that apply*	1=Mechanism application 2=Scope discussion 3=Specific use examples 4=Synthetic literature discussion 5=Implied recommendations 777=No recommendations/can't remember 97=Other, specify
E13.	To your knowledge, which of the following dissemination activities occurred? *Select all that apply*	1=A dissemination event in Washington at USAID headquarters 2=A public dissemination event in the US with government or practitioners 3=A dissemination event in the host country USAID mission office 4=A dissemination event in the host country with government or practitioners 5=The development of a 2 pager or evaluation briefer 6=Dissemination in other venues (e.g., blogs, podcasts, radio) 7=Academic journal article or conference presentation 8=A post-evaluation review and action plan to address whether and how project and evaluation used 9=Dashboards/data visualizations 10=Posted to the in DEC 97=Other, specify 777=None

	Are there any dissemination activities that did not occur, but that you would have liked to see implemented? *Select all that apply*	1=A dissemination event in Washington at USAID headquarters 2=A public dissemination event in the US with government or practitioners 3=A dissemination event in the host country USAID mission office 4=A dissemination event in the host country with government or practitioners 5=The development of a 2 pager or evaluation briefer 6=Dissemination in other venues (e.g., blogs, podcasts, radio) 7=Academic journal article or conference presentation 8=A post-evaluation review and action plan to address whether and how project and evaluation used 9=Dashboards/data visualizations 10=Posted to the in DEC 97=Other, specify 777=None
SECTION F. LOOKING FORWARD *This section is asked to all respondents*		
F1	What advice would you give to USAID, IPs, or evaluators when implementing future IEs?	
F2	I would encourage others at my organization to participate in an IE	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
F3	I would participate in another IE if requested	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree 5=Strongly disagree
F4	I would actively seek out opportunities to participate in another IE	1=Strongly agree 2=Agree 3=Neither agree nor disagree 4=Disagree

		5=Strongly disagree
F5	Which statement best represents your opinion about the future of USAID/DRG support for IEs?	1=The DRG Center should increase its support for IEs 2=The DRG Center should support IEs at their current level 3=The DRG Center should decrease it support for IEs 4=The DRG Center should cease conducting IEs
F6	What advice would you give to the DRG Center on the future of IEs?	*Text*

ANNEX 8. RETROSPECTIVE QUESTION THEMES

RETROSPECTIVE QUESTION	THEME
Q1: Costs	Cost
Q2:Findings	Findings: Coding (Positive, Negative, Null)
Q2:Findings	Findings: Detailed
Q3: Challenge and Lessons Learned	Design - IE clinic or origin
Q3: Challenge and Lessons Learned	Design - Procurement of IP
Q3: Challenge and Lessons Learned	Design - post-procurement
Q3: Challenge and Lessons Learned	Implementation - intervention implementation
Q3: Challenge and Lessons Learned	Implementation - evaluation implementation
Q3: Challenge and Lessons Learned	Implementation - IP-Eval coordination
Q3: Challenge and Lessons Learned	Lessons learned design/procurement
Q3: Challenge and Lessons Learned	Lessons learned Implementation
Q3: Challenge and Lessons Learned	Lesson learned DRG Center
Q3: Challenge and Lessons Learned	Lessons learned Mission
Q3: Challenge and Lessons Learned	Lessons learned: IP
Q3: Challenge and Lessons Learned	Lessons learned: evaluators /PI
Q4: Use	Dissemination
Q4: Use	How it was used
Q4: Use	Why or why not it was useful
Q4: Use	Academic use

RETROSPECTIVE QUESTION	THEME
Q5 Recommendations	Recommendations general
Q5 Recommendations	Recommendations - when to do an IE
Q5 Recommendations	Recommendations - how use IEs more

ANNEX 9. CONJOINT EXPERIMENT PROTOCOL

Because this report is fully retrospective, we could not actively vary components of impact evaluations. As such, lessons learned in this report are based on a select set of DRG IEs that were uniform in key respects, such as in the randomization of treatments, and thus lacked meaningful variation from which we could make other conclusions. In the survey, therefore, we conducted a conjoint survey experiment wherein we asked respondents to evaluate a number of hypothetical impact evaluations and rate which of them they preferred. This enabled us to examine what stakeholders value, in principle, in impact evaluations.

All survey respondents, regardless of stakeholder type, were given five hypothetical IE profiles and asked to choose which of the profiles they preferred. We then conducted average marginal components analysis (AMCE) on the data and present the core results in the manuscript. There we also discuss relevant subgroup effects. We preregistered the expectations and approach with EGAP on March 5, after the experiment was fielded, but before any results were analyzed, and we also submitted IRB clearance through the University of Texas at Austin.

In what follows, we detail the experimental protocol, including the conjoint components, response measurement, and more broadly estimation and analysis. Details about respondents are discussed in the main report.

PROCEDURES

After answering a series of questions about engagement with USAID DRG-relevant impact evaluations, respondents were asked to consider five hypothetical impact evaluation comparisons and then choose which of the five they preferred. The profiles were presented as follows:

CONJOINT PROFILES

The set of impact evaluations carried out by the USAID DRG center lacks variation on some core dimensions. As such, we would like to present some components of hypothetical impact evaluations, which offer more variation on key dimensions of interest, and understand more about which are most useful for USAID learning and utilization. Accordingly, we will now give you dimensions for two hypothetical impact evaluations. Please read each of the descriptions carefully and then provide ratings to three questions that follow. A final question will then ask you to indicate which of the two impact evaluations would lead to the most USAID usage of the results. Please note that we will have you repeat this task 5 times with different profiles.

DIMENSION	IMPACT EVALUATION 1	IMPACT EVALUATION 2
Evaluation approach:	Beneficiaries randomly chosen to receive project	Beneficiaries purposively chosen to receive project
Evaluation results:	Confirm original project expectations	Do not confirm original project expectations
Evaluation conducted:	By an international team with a partner in the field	By an international team
Costs of the evaluation:	Approximately 2% of overall project costs	Approximately 4% of overall project costs
Evaluation results released:	Before future projects are developed	While future projects are in development
Intervention implemented:	As planned at outset without field challenges	With adaptations because of field challenges
Timeliness:	No major delays	Some notable delays

MECHANISM QUESTIONS

On a scale from 1 to 7, please rate each of the impact evaluations.

1. How likely are the results of the impact evaluation to be formally presented within the mission, such as to other project teams or to leadership?
    a. Sliding scale ranging from 1 to 7 for A, where 1 is highly unlikely and 7 is highly likely. (Respondent sets level on the scale)
    b. Sliding scale ranging from 1 to 7 for B, where 1 is highly unlikely and 7 is highly likely. (Respondent sets level on the scale)
2. How likely are the results of the impact evaluation to be formally presented outside of the mission, such as at a results summit?
    a. Sliding scale ranging from 1 to 7 for A, where 1 is highly unlikely and 7 is highly likely. (Respondent sets level on the scale)
    b. Sliding scale ranging from 1 to 7 for B, where 1 is highly unlikely and 7 is highly likely. (Respondent sets level on the scale)

DICHOTOMOUS CHOICE RESPONSE

Which of the impact evaluations would contribute most to USAID usage of the evaluation results?

Choice:

1. Impact Eval A
2. Impact Eval B

ANNEX 10. USAID/DRC INTEGRATED GOVERNANCE ACTIVITY IE FOLLOW-UP

This after action review and analysis was not included in the report because it was not yet completed or close to complete at the time of data collection. Given the rich set of lessons learned from each impact evaluation it is recommended that the E&L team continue to conduct after action reviews at the conclusion of future impact evaluations.

DESCRIPTION

The IE of USAID/DRC Integrated Governance Activity (IGA) took place between 2017 and 2022 for a total cost of $1,518,478.

It was implemented by DAI Global and evaluated by NORC. The IE was designed to measure whether governance interventions integrated with health programming can help improve health outcomes. The IE measured the impacts of three interventions: 1) capacity-building (CB) trainings designed to strengthen the resource management skills of local health service providers and 2) a community scorecard (CSC) intervention aimed at increasing citizens' awareness of health service provision and their ability to mobilize and hold service providers accountable. The third intervention was a combined CB and CSC treatment. The CB intervention trained providers and community health development committee (*comité de développement de l'aire de santé*, CODESA) members; the CSC intervention engaged these groups as well as citizens to assess the performance of health centers and develop community action plans to address shortcomings. These interventions sought to improve health service delivery, perceptions of health services and health governance, health-seeking and health-promoting behaviors, and, ultimately, household and community health outcomes.

FINDINGS

The IE found mixed uptake and implementation results. The CB training was implemented roughly in line with the design, but it did not improve the health centers' staff's ability to answer basic knowledge questions directly related to the training curriculum. The CSC meetings do not appear to have been widely publicized or attended. Moreover, the CB+CSC treatment did not improve CODESA and household survey respondents' ability to answer basic knowledge questions directly related to a component of the intervention. The findings indicate that, in the short and medium terms, neither the CB training alone nor the CB training combined with the CSC intervention substantively improved outcomes in the following categories: health service delivery, citizens' perceptions about health care provision and governance, citizens' health-seeking and health-promoting behaviors, or health outcomes. However, the IE did find some conditions that were more conducive to creating impacts than others, including Has with high social engagement, political efficacy, and horizontal accountability relationships.

CHALLENGES AND LESSONS LEARNED

NORC identified the following challenges during the IE that may have contributed to the null or negative findings.

- Mixed uptake and implementation results. Only 46 percent of the treatment group was found to be compliant at endline. CSC meetings were not widely published or attended.
- Spillover from CB-like or CSC-like activities not implemented by DAI in control communities.
- IE interventions were perceived as "foreign" and allowed for limited local input.

UTILIZATION

At the conclusion of the research, the team developed a findings report to share the results of the IE with USAID/DRG, USAID/DRC, DAI/IGA, and the wider public. The report is available on the DEC. USAID/DRG, USAID/DRC, DAI/IGA, and NORC also participated in a workshop where NORC presented the results of the IE, DAI/IGA presented lessons learned, and all parties discussed possible recommendations. NORC used this discussion to draft the recommendations section of the final report. The baseline, midline, and endline data was also shared with DAI/IGA to allow them to incorporate the information into their own monitoring and evaluation and other reporting efforts.

Though not directly linked to the findings or recommendations of the IE, the DRC government has recognized the scorecard intervention had recognized the scorecard intervention that DAI/IGA and NORC developed for the IE as a best practice and was using it is a model to be implement elsewhere across the country.

RECOMMENDATIONS

On March 22, 2022, the Evidence and Learning Team at USAID's Center of Excellence on DRG facilitated a discussion with USAID/DRC, DAI, and NORC. The following recommendations for future programming and future IEs reflect DAI's exchange of ideas that followed the presentation of the main IE results as well as DAI's comments on the IE.

- Recommendations for future programming:
  - Allow time to define the training needs of participants and design a training program that meets those needs; this may involve adopting multiple approaches to training (e.g., coaching or mentoring and on-the-job training).
  - Target information and mobilization interventions in areas where treatment uptake and effectiveness are more likely (e.g., areas where service providers receive close monitoring from higher-level administrators and communities with higher levels of social engagement).
  - Design and implement activities with meaningful local engagement and buy-in.
  - Expect, plan, and budget for additional logistical and managerial challenges when implementing interventions in hard-to-reach areas.

- Recommendations for future IEs:
  - Consider designing IEs with bundled interventions to increase the chances of producing measurable effects.
  - Weigh the costs and benefits of a panel design in hard-to-reach, insecure areas when designing an IE.
  - Define the roles and responsibilities of USAID, the IP, and the evaluation team from the beginning and sustain collaboration throughout the IE