



# Rigorous Outcome Performance Evaluations (ROPEs) in Democracy, Human Rights, and Governance

**Abstract:** USAID’s Democracy, Human Rights, and Governance (DRG) programming currently relies heavily on two types of evaluations: impact evaluations (IE) and basic performance evaluations (PE), which typically entail around three weeks of largely interview based data collection. However, these two forms of evaluation leave a large “missing middle” and do not adequately meet all DRG evaluation needs. IEs generally provide the greatest degree of confidence assessing program impact, but they are challenging to implement and not the right fit for many interventions. By contrast, basic PEs are very easy to implement, but they are generally not able to accurately measure changes in outcomes over time, much less determine program impact. This document aims to throw a “rope” to practitioners looking for options in this missing middle by introducing what we term Rigorous Outcome Performance Evaluations (ROPEs). These evaluations include data collection at multiple points in time (e.g., baseline, midline, and endline) to measure changes in intended outcomes. This is combined with a process tracing methodology that explores potential intervention contributions by testing progress along the theory of change and considering alternative explanations. Such a design is typically mixed methods and might include quarterly administrative data, multiple waves of survey data, or multiple waves of qualitative data collection. While a promising approach, ROPEs do not have a comparison group that can estimate the counterfactual.

## **August 2022 (Updated February 2024)**

This document was prepared by Daniel Sabet of the Evidence and Learning Team at USAID’s Bureau for Democracy, Human Rights, and Governance (DRG). The document draws on existing DRG resources and new research but is not formal guidance. Please contact [drg.el@usaid.gov](mailto:drg.el@usaid.gov) with questions about this document.

## Introduction

Following guidance from the Office of Management and Budget (OMB), revisions to USAID's Automated Directives System (ADS) in 2020 divided "performance evaluations" into outcome, developmental, formative, and process/implementation evaluations ([ADS 201.3.6.4](#)). Traditionally, USAID performance evaluations (PE) tend to be a one-time study, involving three to four weeks of mostly qualitative field work, what we term a basic PE. These one-shot evaluations have difficulty measuring outcomes and measuring changes in outcomes over time, as evaluators must reconstruct and estimate how changes have or have not occurred. What then should a USAID DRG outcome performance evaluation design look like? In this document, we outline the case for what we call Rigorous Outcome Performance Evaluations (ROPEs). These are longer-term evaluations, or longitudinal designs, that measure outcomes at multiple points in time, combined with a process tracing methodology. This design offers at least three advantages over the more common basic PE:

1. By comparing baseline, midline, and endline data, longitudinal PEs can measure changes in outcomes over time with far greater confidence.<sup>1</sup>
2. Baseline and midline data and analysis can inform the intervention prior to a summative endline evaluation.
3. While still limited in attributing changes to an intervention, a long-term PE is better suited to consider possible intervention contributions by exploring progress along the theory of change and alternative explanations for changes in outcomes.

## An option for examining program outcomes when an IE is not the right fit

USAID prefers the use of impact evaluations (IEs) to test if an intervention is achieving a specific outcome ([ADS 201.3.1.3](#)). However, there might be a variety of reasons why an IE is not the best fit methodology, for example: it might not be possible to identify a control group; there might be insufficient treatment units; the costs and challenges of an IE might outweigh the benefits; there might be ethical reasons not to randomize treatment; and/or a new intervention approach might be insufficiently consolidated and require extensive adaptation before being tested with an IE design.

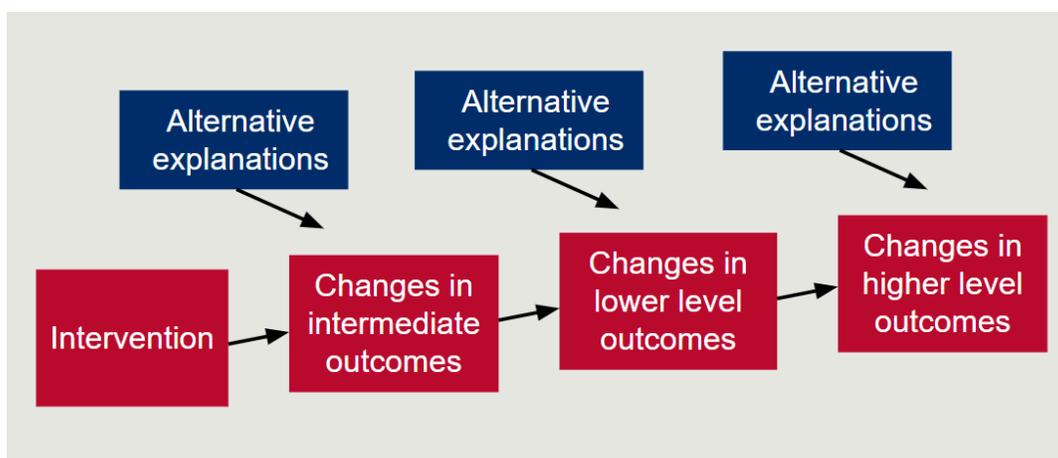
While the longitudinal design discussed here does not compare treatment groups with a control group, evaluators can still consider alternative explanations or hypotheses that might explain why we do or do not observe changes over time. This might be done through statistical control (e.g., regression analysis) or qualitative methods to eliminate alternative explanations. For example, a common method to

---

<sup>1</sup> Long-term evaluations might also provide more accurate measurements, particularly for qualitative data. DRG work is inherently political work that often risks window-dressing reforms or de jure changes rather than meaningful reform. These challenges create an enormous measurement problem for basic, one-time performance evaluations. There are often strong incentives for interviewees to speak less than frankly, overstate problems under political predecessors, and exaggerate recent accomplishments. While good evaluators are sensitive to these challenges, they often must make sense of contradictory interview evidence and struggle to understand the whole picture in one data collection trip. Multiple trips, multiple interviews with the same individuals over time, and incorporation of complementary data sources allows evaluators to improve their measurement accuracy and confidence in their conclusions.

consider program contributions is process tracing, which is clearly explained by Joe Amick in a [three-part module for USAID](#). Process tracing is a good fit for evaluating many USAID interventions because it builds off a program’s theory of change (TOC). The method entails clearly laying out the TOC, measuring outputs and outcomes at each step in the theory, and considering alternative explanations for changes besides the intervention (See Figure 1). Process tracing is a form of contribution analysis, which means that it does not tell us the precise impact of an intervention on an outcome, but it is able to show if an intervention contributed to an outcome or not.<sup>2</sup> Perhaps more importantly for USAID, as Amick notes, it can show where a TOC might be incorrect, where a causal link breaks down, or what complementary factors increase or decrease the effectiveness of an intervention.<sup>2</sup>

*Figure 1: Traditional process-tracing approach to explore the contribution of a program*



### When to use a ROPE

1. When users want more accurate measures of changes in intended outcomes over time.
2. When users are most interested in project outcomes and impact but unable to conduct an IE.
3. When outcomes are not likely to improve on their own or primarily because of other identifiable factors that would make it hard to determine USAID contribution.

### Examples

Consider the following illustrative examples of longitudinal performance evaluations of DRG interventions:

- **Juvenile justice reform in Guyana, St. Lucia, and St. Kitts and Nevis (USAID):** To assess reform efforts to divert youth from the juvenile justice system, the evaluation team conducted a qualitative study at three points in time over a five-year period to measure reform progress. During each study wave, the evaluation team reviewed changes in policies and legislation, conducted interviews with key sector stakeholders and conducted interviews and

<sup>2</sup> Variants of this approach include [contribution analysis](#), [causal link monitoring](#), [qualitative impact assessment protocol](#).

focus groups with youth in the juvenile justice system.<sup>3</sup>

- **Reform of the electrical utility in Malawi (MCC):** Through the use of process mapping, the evaluation drilled deep into eleven different electrical utility processes expected to change as a result of the intervention (e.g., connecting customers, responding to outages, annual budgeting) at three points in time. Through the process mapping approach, the evaluation team was able to identify quantitative indicators and administrative data to assess if change had occurred and use qualitative data to consider the Millennium Challenge Corporation's (MCC) contribution.<sup>4</sup>
- **Procurement reform in Honduras (MCC):** To assess the contribution of an MCC procurement reform initiative, the evaluation team conducted a baseline and endline survey of vendors who sell goods and services to the government to measure changes over time. They then used qualitative data from audit reports and interviews to consider the role of the intervention in observed changes. The qualitative data also explored process level changes along the project's theory of change.<sup>5</sup>

As noted in the descriptions, all three evaluations were mixed methods, they relied heavily on qualitative data, but they complemented this with administrative and survey data. In all three cases, the evaluation was able to (1) confidently document the change or lack of change over time, (2) consider the influence of alternative factors, (3) understand and document the obstacles along the theory of change, (4) make reasonable conclusions about the performance of the intervention, and (5) make recommendations to inform the intervention at baseline and midline.

## Challenges

Like other approaches, a ROPE is not without its challenges. The following paragraphs first explore the challenge in assigning attribution and then discuss implementation challenges.

- **Challenge in assigning attribution:** While offering a better measure of change over time, as noted above, a longitudinal design still confronts a major limitation in attributing those changes to an intervention. Other factors that vary over time, such as changes in the political landscape, the economy, or other donor or government initiatives might offer better explanations for observed changes than an intervention. This is the principal reason why evaluators typically prefer randomly-assigned treatment and control groups in which the only difference between

---

<sup>3</sup> Barnes-Ceeney, Kevin and Lily Hoffman (2020) [Evaluation of Juvenile Justice Sector Reform Implementation in St. Lucia, St. Kitts and Nevis, and Guyana](#). USAID. Another robust longitudinal PE conducted under DRG-LER is the [Endline Performance Evaluation of USAID/Displaced Children and Orphans Fund \(DCOF\)'s Family Care Project in Burundi](#).

<sup>4</sup> Sabet, Daniel, Meredith Feenstra, Arvid Kruze, and Linda Chinangwa. (2019) [Malawi Power Sector Reform Project: Final Endline Performance Evaluation Report](#). Social Impact.

<sup>5</sup> Sabet, Daniel, Nick Livingston, Albert Pijuan, Irma Romero, Lisette Anzoategui, and Jordan Sticklin. (2020). [Honduras Threshold Program: Endline Evaluation Report](#). Social Impact.

the two groups is the intervention. To offer an example, a recent IE of an intervention to reduce youth violence risk factors found dramatic reductions in risk factors among the treatment youth; however, they found similar reductions in a control group.<sup>6</sup> Without the control group, the evaluators would have concluded that the intervention was a major success. Tools such as process tracing, identified above, can help address the limitations of not having a control group, but it is important to be clear that process tracing is not a replacement for an IE.

- **Evaluators need the right skill set and availability over time:** ROPEs require skilled evaluators who are comfortable in mixed methods, possess subject matter expertise, and are able to make a long-term commitment to the evaluation. The above-mentioned juvenile justice evaluation had considerable turnover in team leaders, which undermined the benefits of a long-term evaluation approach.
- **Continued obstacles to use:** One of the clear advantages of the longitudinal approach is that baseline and midline evaluation data and analysis offer learning opportunities that can benefit the intervention. Nonetheless, in practice, many implementers remain suspicious of external evaluations and prefer to rely on their own internally-generated analysis. In addition, the slow generation of evaluation reports undermines timely use of information. As such, use will continue to depend on USAID resolve to ensure timely availability of information and formal and informal efforts to build evaluation buy-in and promote use.<sup>7</sup>
- **Greater level of effort, cost, and longer timelines than a basic PE:** ROPEs require a process that looks more like an IE, with significant costs, level of effort, and time requirements. ROPEs might include a scoping trip to design the evaluation, a longer review of data collection instruments, procurement of a survey data collection firm, and more complex data analysis. As such, ROPEs (like IEs) require early planning and budgeting at the design stage of an activity. Nonetheless, the return on that investment of time, effort, and funding can be much greater confidence in the evaluation results than with a basic PE.

## **A ROPE and CLA**

On the one hand, a ROPE and collaborating, learning, and adapting (CLA) go well together as baseline and midline data and analysis can inform adaptation. On the other hand, there is a potential tension

---

<sup>6</sup> Diaz-Cayeros, Alberto, Stephanie Gimenez Stahlberg, Rachel Pizatella-Haswell, Daniel Sabet, and Julia Kresky (2020) [Evaluation of Secondary Prevention in the Community, Family, and Youth Resilience Program in St. Lucia, St. Kitts and Nevis, and Guyana: Final Endline Report](#). USAID.

<sup>7</sup> For example, USAID might request interim deliverables with preliminary findings upon which to base decision making while final deliverables are under development. The evaluation team might participate in a workshop with implementers to discuss baseline and midline findings, and should develop a post-evaluation action plan ([ADS 201.3.6.10](#)) at each stage of the evaluation to respond to the findings, conclusions, and recommendations.

between a ROPE and CLA if adaptation alters outcome variables measured in the baseline. For this reason, it is important to anticipate the extent of adaptation that is possible in the course of the intervention.

- If the high-level outcomes might change, then a ROPE is probably not the best approach, and a developmental evaluation would be a better fit.
- If the high-level outcomes are fixed but the intervention is expected to be highly adaptive (i.e., altering the theory of change and intermediate outcomes) then strict process tracing will be difficult and a more flexible and qualitative approach to assessing contribution will be required.
- Where adaptation is not expected to shift high- or intermediate-level intended outcomes, then a more quantitative and mixed-methods approach is appropriate.

### **Evaluation Scopes of Work**

**Overall approach:** There are two main approaches to a ROPE Scope of Work (SOW). Under the first, there is relative certainty about the evaluation approach, this is reflected in the SOW, and the learning partner is issued a tasking for the whole evaluation. Under the second, there is greater uncertainty about the evaluation approach and budget implications, and learning partners are first tasked with coming up with a design and then, upon approval, tasked with carrying out the evaluation.

**Purpose, users, intended use, and focus:** Operating units should have a clear use-case for the evaluation and this should be reflected in the SOW. It is important to note that while some ROPES focus on the entire intervention over the entire period of performance, others focus on a particular activity or a focused round of implementation.

**Evaluation questions:** There are several different types of questions that are suitable for a ROPE.

- **Changes over time:** Do we see changes over time (positive or negative) in the [core outcomes]?
- **Contribution:** Does the available evidence suggest that the intervention had an effect on [core outcomes]? Why or why not?
- **Theory of change:** Does the intervention's theory of change appear to be borne out by the available evidence? Are there certain assumptions that did not hold? (Such questions are particularly amenable to the process tracing approaches mentioned above.)
- **Explanation:** What factors explain variation in the [outcomes of interest] across [beneficiaries]? (For example, if an intervention is designed to promote youth political engagement, the evaluation would use statistical analysis to explore why some youth participate and others do not.)

**The ROPE process and timelines:** The process can vary depending on the complexity of the evaluation. A good design process typically requires five to six months and preparation for baseline

data collection will require an additional one to five months depending on if a local data collection firm needs to be procured.

ROPEs should be planned at the activity design stage. Evaluators should be procured before (or at the same time) as implementers and the scoping trip and evaluation design process should coincide with implementer start-up and entail extensive evaluator-implementer coordination.<sup>8</sup> This ensures that the evaluation is based on the actual intervention while giving sufficient time to collect baseline data. It is important to keep in mind that even after a scoping trip, it takes several months to finalize a design and procure (if necessary) local data collection firm support. With good coordination, baseline data collection can occur before implementation efforts ramp up and begin to have plausible effects on outcomes.

---

<sup>8</sup> A recent retrospective of DRG IEs offers several lessons learned about evaluator-implementer coordination that also apply to long-term PEs. Michael G. Findley, Aleta Starosta, and Daniel Sabet (2021) [DRG Impact Evaluation Retrospective: Learning from Three Generations of Impact Evaluations](#). USAID.