

UNBUNDLING ACCOUNTABILITY & IMPACT FOR THE NEXT GENERATION OF U.S. INTERNATIONAL DEVELOPMENT EFFORTS

Case by: Christopher Grady, 2026

Table of Contents

/	Summary	1
/	Learning Case	2
/	Conclusion	4
/	Recommended Reading	5
/	Notes	6
/	References	7

Summary

Throughout its 64 year history, the U.S. Agency for International Development (USAID) supported over 100 countries and improved millions of lives. But its functioning was not perfect, and the architecture of international development deserved an overhaul based on those six and a half decades of experience. As a senior advisor at USAID from 2020–2025, I focused on the design and evaluation of the agency’s governance interventions. Though [USAID was working towards evidence-based policymaking](#), its system for monitoring, evaluation, and learning (MEL) still prioritized bureaucratic compliance over assessing impact.

USAID’s rapid shutdown in early 2025 caused unnecessary suffering by throwing the proverbial baby out with the bathwater, but it also offers us an opportunity to reassess many aspects of the international development infrastructure because there is a pause to much of the work. This article offers my reflections on how a next generation of aid (“nextgen AID”)—including the current U.S. model of foreign assistance through the State Department—could learn from the past and create a more effective future for international development.¹

Key Takeways:

- / **Characterizing the output trap:** I briefly describe the previous USAID system for monitoring, evaluating, and learning from its development programs. This system generally revolved around tracking program **outputs**—things the program did—at the expense of assessing our effect on **outcomes**—things the program aimed to change in the world.
- / **Proposing a disaggregated MEL system:** I suggest a system that unbundles **tools for accountability, program management, and portfolio strategy**. By separating these functions, the system can provide the people and organizations implementing development programs (“implementers”) with the management information they need to achieve a program’s goals, while giving funders the strategic information they need to maximize the effect of their overall development portfolios.

Learning Case

USAID's Monitoring, Evaluation, and Learning (MEL) System

USAID traditionally operated by setting global and national development goals and then contracting international or local organizations to implement programs aimed at achieving those goals.² To support portfolios and programs, USAID employed a **Monitoring, Evaluation, and Learning (MEL)** system to serve three distinct objectives: **accountability, strategy, and management**.

- / **Monitoring supports accountability:** “Did they do what we paid them to do?” Monitoring solves the principal-agent problem, ensuring implementers deliver the outputs required by their contracts.
- / **Evaluation supports portfolio strategy:** “Are we funding the right things?” Evaluations provide decision-makers with the information they need to decide which programs successfully impact development outcomes.
- / **Learning supports program management:** “How can this program be better?” Learning helps implementers improve their programs, maximizing the chance that they achieve their development goals.

The main tool USAID used to achieve these goals was called an **Activity Monitoring, Evaluation, and Learning Plan (AMELP)**.³ The AMELP detailed how the implementer would *monitor* what the program does, *evaluate* its effects, and *learn* to improve their implementation. While all three goals are vital, they could not all be satisfied by one system. Because AMELPs first and foremost attempted to solve the principal-agent problem, they became an accountability tool heavily focused on monitoring outputs. This created an output trap and inadvertently incentivized “check-the-box” behavior, where success was determined by what was done (training an adequate number of people or funding an adequate number of businesses), rather than what was achieved.⁴ AMELPs ensured that work was done, but not necessarily that the *right* work was done.

In a tacit acknowledgement that AMELPs were not sufficient for all MEL needs, USAID often commissioned separate **evaluations** to determine the effect of the high-priority programs. However, evaluations were typically plagued by at least one of two problems. First, due to

Learning Objectives

Reflect on the tension between completion (checking boxes by finishing tasks) versus achievement (meaningful learning and success) in evaluating international development program outcomes.

Contextualize USAID's approach to supporting accountability, portfolio strategy, and program management.

Reflect on the limitations of an integrated Monitoring, Evaluation, and Learning (MEL) system.

Consider the benefits of a disaggregated system that features separate evaluation of portfolios and programs, impact-focused evaluation of all new programs, iterative learning, and randomized audits.



political and operational reasons, many evaluations became co-opted for implementer learning (how to improve the program) rather than program evaluation (whether the program worked).⁵ Second, due to insufficient theories connecting individual programs to USAID’s portfolio-level development goals, evaluations often assessed a program’s effect on portfolio-level development outcomes that an individual program had little hope of impacting. This issue led to false negatives: program evaluations would find that a program had no effect on the outcomes it analyzed, though the program may have successfully impacted lower-level outcomes that were not analyzed. In one example from my time at USAID, we evaluated the effect of a program on country-level rule of law scores. We found no effect on that high-level outcome, though the program helped connect thousands of citizens to pro bono judicial and legal assistance.

This system of AMELPs and evaluations failed to serve the needs of both USAID and implementers. USAID lacked the means to understand the effect of its programs and portfolios, and implementers were judged on their ability to move high-level outcomes that could only be affected by an entire development portfolio.⁶

New Possibilities

For a nextgen AID to achieve its development goals, it should unbundle its MEL functions. Instead of a combined MEL system trying to serve three masters, nextgen AID should adopt a disaggregated system that separates tasks for portfolio strategy, program management, and accountability.

The system must support portfolio strategy by providing nextgen AID with holistic and accurate information about how each program contributes to their overall goals. For implementers, the system must hold them accountable for achieving reasonable and discrete program goals, while giving them the freedom to learn how best to implement programs likely to achieve those goals.

For portfolio strategy, nextgen AID should conduct portfolio-level impact evaluations. By articulating high-level goals and assessing the combined effect of a portfolio of programmatic efforts to achieve those goals, nextgen AID holds *itself* accountable for the high-level goals they want to achieve, because they would be able to track if they are keeping their stated portfolio-level goals. Keeping portfolio-level evaluations separate from the program-level evaluations prevents the “false negative” problem by ensuring individual programs aren’t blamed for failing to move entire national indices.⁷

Nextgen AID should also conduct impact evaluations of all new programs. These evaluations should focus on discrete, achievable outcomes that flow from the high-level portfolio goals. They should be agnostic as to *how* results are achieved; just as programs should not be blamed for failing to achieve portfolio-level goals beyond their control, they should not be credited for imple-

Discussion Questions

How are the challenges presented by USAID’s MEL system unique to international development? How are they similar to MEL systems in other government agencies, such as the U.S. Department of Agriculture or the U.S. Department of Energy?

Can you think of another international development example where a program was evaluated on portfolio-level outcomes instead of program-specific priorities?



menting ineffective outputs. *Impact* evaluations should not inherently care about the number of meetings held or trainings conducted, because the implementation of activities is not itself evidence of a program's impact. By evaluating whether outcomes were achieved rather than whether implementation followed a pre-decided process, nextgen AID can shift the incentive for implementers away from “completing outputs” and towards “improving outcomes.”

For program management, decouple learning from compliance. nextgen AID should allow implementers to control their own learning. Rather than pre-specifying outputs, allow implementers the freedom to pilot different output activities and identify which are most effective.⁸ Then conduct an impact evaluation on the final program the implementer designs, not the activities they pilot during the iterative learning process. This freedom restores the “L” in MEL as a tool for learning rather than a record of bureaucratic deviation.

For accountability and fraud prevention, adopt randomized audits. To maintain fiscal integrity without stifling innovation, nextgen AID should adopt a “trust but verify” model. Implementers report their outputs, and nextgen AID randomly audits a small percentage to prevent fraud. If fraud is detected, the consequences should be severe (program cancellation and financial repayment). If fraud is not detected, nextgen AID can combine output information with their impact evaluation to understand what types of outputs do/do not contribute to the program's discrete goal.⁹

How would this work in practice? Imagine nextgen AID's national-level goal is greater democratic accountability in a country, and they fund multiple programs that contribute to that goal.

- / **First**, nextgen AID should evaluate the effect of their portfolio on democratic accountability.
- / **Second**, if one of the programs is meant to contribute to democratic accountability by increasing voter turnout, a key component of holding governments accountable to citizens, that program should be evaluated on how it impacts voter turnout.
- / **Third**, the implementer can pilot several ways to improve voter turnout and select the best method (or combination of methods) to do so.
- / **Fourth**, if that program reports that they funded voter registration centers to increase voter turnout, nextgen AID should select 10% of those registration centers to ensure that they were, in fact, funded.

Conclusion

The future of U.S. international development should resist the inertia to simply rehash old systems. International development practitioners have a rare opportunity to reflect on and improve the system for monitoring, evaluating, and learning from programs, moving it away from a focus

Discussion Question

What are some of the potential benefits and drawbacks of a disaggregated MEL system?

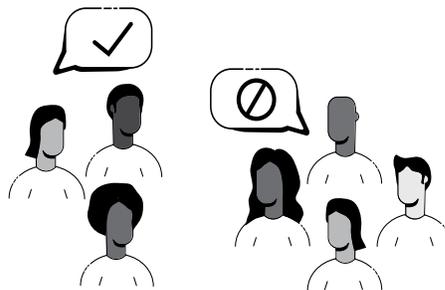


on bureaucratic output compliance and towards a focus on development outcomes. This proposed system unbundles processes for accountability, strategy, and management so that they are no longer working at cross-purposes.

Much attention has been paid to [fixing USAID program evaluations](#), but fixing program evaluations is not enough. To improve the lives of people around the world, nextgen AID needs to learn about the impact of its portfolios, not just individual programs. Nextgen AID should thus begin with rigorous evaluation plans for their development portfolios, even before getting into the details of individual programs.

But the individual programs cannot be ignored. Nextgen AID should also evaluate individual programs based on their impact of discrete, achievable program-level goals. By focusing program evaluations on discrete outcomes rather than heavily-monitored outputs, we create the space for what Dan Honig calls [navigation by judgment](#). In complex development environments, success often depends on the ability of implementers to use their local expertise and “soft information” to pivot when a plan isn’t working. When we stop measuring a program by how many brochures it prints and start measuring it by the outcomes it achieves, we give implementers the freedom and incentives to achieve those outcomes.

A disaggregated MEL system does three things: (1) allows nextgen AID to assess if and how its portfolio of programs achieves higher-level goals, (2) provides incentives for implementers to achieve development goals rather than deliver specific outputs, and (3) frees professionals to use the judgment necessary to solve the world’s most difficult development challenges. The success of U.S. foreign assistance must be measured not by the volume of work it performs, but by the tangible change it leaves behind in the world. The next generation of U.S. foreign assistance must have a MEL system up to this task.



Recommended Reading

- / [Actually Navigating by Judgment: Towards a New Paradigm of Donor Accountability Where the Current System Doesn't Work](#), by Dan Honig (2020) for the Center for Global Development, Washington, D.C.
- / [It's All About MeE: Using Structured Experiential Learning \('e'\) to Crawl the Design Space](#), by Lant Pritchett, Salimah Samji, and Jeffrey S. Hammer (2013), available at SSRN.
- / [Escaping capability traps through problem driven iterative adaptation \(PDIA\)](#), by Matt Andrews, Lant Pritchett, and Michael Woolcock (2013), published in *World Development*.

Notes

1. My experience was with USAID's Democracy, Rights, and Governance (DRG) programs. The MEL processes I discuss apply to other USAID sectors, but the evaluation challenges are particularly poignant for DRG because DRG outputs are further from our DRG outcomes than outputs and outcomes from other sectors. For example, a program to train journalists (DRG output) is less closely tied to an improved information environment (DRG outcome) than administering vaccines (health outputs) are to disease immunity (health outcomes).
2. When I refer to "USAID" in this article, I refer to the many different bureaucratic units within USAID. USAID is a large organization and referring to it as a single entity is an oversimplification, but the large gain in parsimony is worth the minimal loss of precision.
3. USAID's Operating Units (i.e., Missions and Bureaus) would monitor their portfolios with different tools, like Performance Management Plan (PMPs). They share some of the same issues as AMELPs, but since this document focuses on evaluation at the program level, those tools are beyond the scope of this article.
4. This phenomenon has been called [isomorphic mimicry](#), where organizations focus on appearing successful through procedural compliance while the actual functional impact is lost. The deeper reason for this focus on output compliance is that USAID, like most government agencies, was held accountable by Congress for what was done with their money, not for getting results. This leads to incentives that ensure accountability but not impact.
5. The use of evaluations came after an influential 2008 [report from the National Research Council](#) recommended that USAID adopt impact evaluation methodologies, especially randomized controlled trials (RCTs), for evaluating the effects of their programs. However, impact evaluations were almost never used in many development sectors. Based on our internal data, USAID completed 213 evaluations in the DRG sector alone between 2016 and 2022, but only 19 were impact evaluations.
6. The question of why the USAID MEL system functioned this way is beyond the scope of this article.
7. Portfolio-level evaluations should be considered evaluations of the development strategy as a whole, not a particular institution. Development portfolios are constrained by earmarks and (mutable) political priorities, so even a development agency like USAID had limited control over its goals and the programs it could implement to achieve those goals.
8. The details and length of the pilot period will likely vary by program.
9. It is important to note that fraud was incredibly rare for USAID programs.

References

- / Andrews, M., Pritchett, L., & Woolcock, M. (2013). Escaping capability traps through problem driven iterative adaptation (PDIA). *World Development*, 51, 234-244. <https://doi.org/10.1016/j.worlddev.2013.05.011>
- / Honig, D. (2018). *Navigation by judgment: Why and when top-down management of foreign aid doesn't work*. Oxford University Press.
- / Rose, S. (2020). Establishing USAID as a leader in evidence-based foreign aid. Washington, DC: Center for Global Development. <https://www.cgdev.org/publication/establishing-us-aid-leader-evidence-based-foreign-aid>.
- / National Research Council, Policy, Global Affairs, Security, Cooperation, & Committee on Evaluation of USAID Democracy Assistance Programs. (2008). *Improving democracy assistance: Building knowledge through evaluations and research*. National Academies Press.