The ELMA Philanthropies
Measurement & Evaluation Principles and Practices

June Lee
Director, Measurement & Evaluation

Purpose and Audience

Measurement and Evaluation (M&E) at ELMA is guided by a set of principles that aim to contribute to effective, impactful philanthropic investments for The ELMA Group of Foundations. These M&E principles have been informed by multiple sources: ELMA's overarching approach to philanthropy, formal and informal input from staff (program and non-program) and board members,[1] and an extensive review of other foundations' M&E thinking and practice.

This document is intended for Philanthropies staff and our evaluation partners. It codifies a set of shared principles and guidance for M&E, with the ultimate goal of yielding objective data that assess the social returns on our investment, and can contribute to evidence bases that may inform broader policy, practice, and funding. It does not explore all the technical and non-technical considerations that should go into an evaluation.

A Note on "Measurement" vs. "Evaluation" vs. "Monitoring"

Different terms are used to refer to the measurement and evaluation function at foundations (e.g., learning, evaluation, evidence, strategic impact, research). Measurement is the broadest definition of activities that assess and/or quantify the progress or outcomes of ELMA's investments. It can include data from commissioned evaluations, investments' measures of success, and grants management. All are used to generate evidence—of varying depth—that grant investments are on the path to success. Evaluation and monitoring are measurement activities and are complementary (see Table 1). Monitoring is the ongoing tracking of programs to ensure that they stay on course, and typically refers to routine activities conducted by implementers/grantees. ELMA staff also actively monitor grant investments—through our interim reporting processes, formal and informal check-ins with grantees, site visits, etc.—to assess if they are on or off track in reaching their goals, and to provide additional support as necessary.

Evaluations can take many forms; they are distinct from monitoring in that they aim to provide an objective, systematic, and independent appraisal of the program (e.g., through a deeper dive into elements of a program model, assessing theories of change). A range of methods can be used in evaluations, including experimental, quasi-experimental, process, and qualitative methodologies. They can focus on processes or outcomes, and answer broader questions about lessons learned. Evaluations are usually (but not always) conducted by an independent third party and can be commissioned by ELMA, a co-funder, or the grantee. ELMA M&E staff oversees ELMA-commissioned evaluations and in the case of co-funder or grantee-commissioned evaluations, provide input to shape them. Data from both monitoring and evaluation can be used to validate an investment's measures of success.

The intended user of M&E results can be the implementer/grantee; ELMA; governments to make, fund, or fundraise for a policy change; other funders when making funding decisions related to a specific organization or intervention; other implementers who want to test or adapt an intervention; or the wider world (e.g., academic research audiences). The type of evaluation and approach can depend heavily on the evaluation's primary user.

[1] This included a survey of board, program, and non-program staff involved in grant recommendations was fielded in March 2018 to understand staff and board perspectives on the priorities and gaps in ELMA's current M&E practice.

Table 1. Monitoring vs. evaluation

	Monitoring	Evaluation
Purpose	• Focus on implementation, comparing what is being implemented vs. what was planned. • Keep project on track; identify areas of underachievement; make course corrections • Provides lessons learned and offers recommendations	• Can focus on process and/or outcomes • Validates what was achieved • Examines how and why outcomes were (or were not) achieved • Examines issues such as best practices, adaptation of models, scale, sustainability • Provides lessons learned and offers recommendations
Timing	Routine, regular, throughout the project	• At significant points during or after the project • Selective and episodic
Scope	• Activities/inputs • Outputs • Outcomes	• Process, intermediate or final outcomes • Test underlying theory of change (or parts thereof) or dive deeper in elements of a program's model • May try to understand causality
Sources/ Participants	Internal staff or implementers	Internal and external: Evaluators, implementers, funders, stakeholders, partners
Users	Internal staff or implementers Funder	Internal and external: Implementers, funders, stakeholders, partners
Examples	Routine monitoring data on reach, training, convenings, etc.; administrative data	• Needs assessment • Process/implementation evaluations • Outcome evaluations

Table 1 describes the attributes typical of monitoring and evaluation, with some characteristics that are common to both (e.g., monitoring and process evaluations can yield similar information).

This document focuses largely on evaluation (rather than monitoring) and outlines five interconnected M&E principles, and their implications for practice.

1. Evaluations should be forward-looking and prioritize decision making and learning

ELMA engages in evaluation activities to assess the effectiveness of program strategies and specific investments, inform internal decisions, and inform and ideally influence other actors' decisions. Decision making and learning[2]—rather than accountability or policing—is the core focus of M&E and **evaluations should be linked to decision points so that they are timely and useful.** A common pitfall in evaluations is that findings are designed and delivered too late to be useful. For example, an impact evaluation of a two-year program may be completed at the end of the third year, and ELMA has had to make investment decisions in absence of the evaluation findings. Evaluations are of limited utility at that point, and intentions to "keep the results in mind for future investments" are often unfulfilled. In some cases, program outcomes are not fully realized until some time after the end of the grant, and outcomes data are not yet available for decision-making, in which case information from other ongoing

[2] "Learning" in this case prioritizes learning on the part of ELMA, implementing partners, and/or policymakers, rather than academic learning.

M&E will help to validate that the program is on the path to achieving impact and provide confidence in investment decisions.

Evaluations also cannot inform decision making without a clear plan for their use. Therefore, measurement and evaluation planning must happen early in the grant development and strategy-setting process, and M&E should be planned in tandem with program design.

Once it is determined that a grant or cluster can benefit from external evaluation, the plans must be designed with a clear understanding of:

1. what decisions need to be made
2. when they need to be made
3. what specific information is needed to make the decisions
4. what ELMA would do differently if we had that information
5. who (internal and external) will use that information

If a program aims to influence policy or practice, additional considerations include:

6. who the audience for the evaluation is
7. how the evaluation fits within a broader influence or advocacy strategy

The evaluation planning tool in Appendix A is intended to facilitate these discussions.

It is important to note that results from evaluations are one piece—but not the only piece—of information used in ELMA's decision-making.

2. We prioritize evaluations for grant investments that are large and/or otherwise strategically important and where evaluation results have the potential to influence policy, funding, and practice

Engaging external evaluators to formally evaluate every program in which ELMA invests is not a feasible use of time or resources. We keep several criteria in mind when thinking about whether a grant or cluster can benefit from an evaluation, what kind, and how. These considerations fall along the dimensions of importance and influence (see Fig. 1).

Importance indicates the strategic importance of the grant to ELMA, including:
1. expected outcome/impact
2. size of grant investment[3]
3. ELMA's long-term commitment to the strategy
4. a program's potential scalability
5. a program's potential to improve policy or practice

Influence denotes the potential for evaluation findings to have influence (especially externally), such as:
1. potential learning (e.g., new and/or untested approaches; plausibility of grantee's theory of change; contribution to the evidence base)
2. urgency for course corrections or future funding decisions
3. potential leveraging (e.g., providing a basis for securing additional funding)
4. alignment with grantee and stakeholder/government concerns
5. a program's potential to improve policy or practice

[3] Not every strategically important grant is large, but large grants are by nature strategically important because of the substantial investment that ELMA has made.

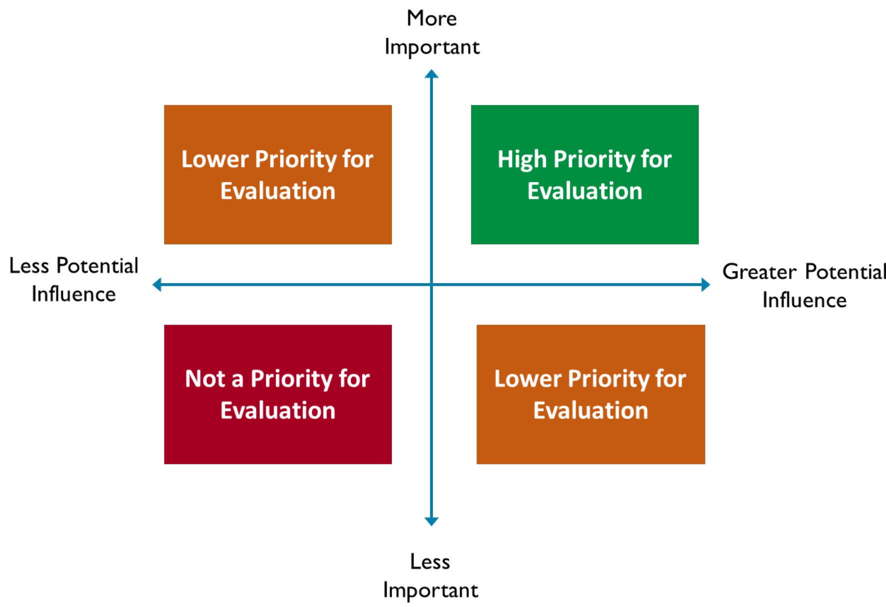

Figure 1. *Weighing importance and influence to prioritize evaluation*

Not all criteria receive equal weight, and different criteria may be used to deliberate what to evaluate versus how intensely to evaluate. The more a grant or cluster meets the criteria above, the more important it is to engage in serious and thoughtful planning for evaluation. Appendix B contains a checklist of considerations to help prioritize evaluations.

3. Evaluations are fit for purpose and right-sized

Fit for purpose: The purpose of the evaluation determines its design.

Evaluations should be fit for purpose. That is, the purpose of the evaluation should determine its design. (see Table 2). Different kinds of evaluations can help to distinguish between failures of theory (i.e., program strategies do not lead to the intended outcome because the Theory of Change is flawed) and failures of implementation (i.e., the theory may be correct, but resources and capacities were insufficient to implement it and produce results), barring unexpected external shocks/factors. Understanding when each type of evaluation is useful and what kind of information is needed will help to ensure evaluations serve their purpose and are designed accordingly. The evaluations described below can be used to inform a program before or during implementation, or assess processes and outcomes after the fact.

Table 2. Types of evaluation

Type of Evaluation	Purpose	When?
Needs assessment	Determines the need for the intervention among a specific population, and how/if the intervention can meet the need; assesses where and how a program should work	Developing new program or modifying a program for a new context/population

Type of Evaluation	Purpose	When?
Process/Implementation Evaluation	Examines whether a program is implemented as designed among the target population. Identifies barriers and enabling factors to implementation. Enables early course correction Often complemented by monitoring data	The onset of a program and on an ongoing basis
	Look more deeply at issues that arise during the course of implementation	During or after program implementation
	Lessons from implementation process that guide continued movement to scale and/or inform other efforts	After implementation/at the end of the program
Outcome evaluation	Examines prospects for achieving outcomes, leading indicators, and changed assumptions	Early or during a program
	Measures medium- and long-term outcomes from a program to understand whether a program achieved its objectives, for whom, and how. Outcomes can happen long after the end of a grant. Assess the strength of a causal link Can include cost-effectiveness analyses	During, at the end of, or after the end of a program
	Understand how impact is sustained, and (where relevant) contribution to systemic change	At or after the end of a program

In addition to thinking about what type of evidence is necessary, we also consider the *level* of evidence needed. The standard of evidence should be proportional to the nature of the investment and the decisions or actions to be made. For example, a program that we hypothesize can or should be scaled widely or that has implications for practice and policy prescriptions likely requires an established, rigorous evidence base of its effectiveness and scalability. In fact, the evidence generated can contribute to more funding and effective scaling. M&E for programs in a phase of testing and adaptation, on the other hand, may be oriented toward learning and establishing sound measurement practices rather than on rigorous proof of impact. Decisions around every evaluation have to balance among factors such as technical considerations, cost, time, operations, political considerations, etc. A consultative process that enables rich discussions that weigh different dimensions of rigor and practicality will help to achieve the optimal M&E design. As mentioned earlier, the type of evaluation and approach can also depend heavily on the evaluation's primary user.

Different quantitative and qualitative evaluation methods are complements to improving understanding, and we believe in their judicious use. Some evaluation designs (such as randomized controlled trials) can be complex and costly and have many design requirements. When it is impossible to meet these requirements or where key aspects of evaluation design have to be compromised, resources can be put to better use by bolstering the quality of monitoring data.[4]

Right-sized: How much should an evaluation cost?

As described above, the cost of an evaluation can vary widely depending on its goal and design. A survey of U.S. foundations revealed that evaluations comprised 3.7% of foundations' programmatic budgets on average.[5] Smaller foundations spent a larger share of their budgets on evaluations because the costs of evaluations do not rise proportionately with program costs. Conventional rules of thumb have pegged evaluation budgets at 5-10% of programmatic budgets, though rigorous experiments fall outside the range. The Scaling-Up Communities of Practice M&E working group advocates allocating as much as 20% of the budgets of pilot projects to monitoring, evaluation, learning, and communications. Rather than relying on arbitrary prescriptions, however, we believe that the cost of an evaluation should be proportionate to its purpose, design, importance of decisions it informs, and the value of evidence it generates.

4. M&E is collaborative and consultative

For evaluations be useful, they must be designed in consultation with their primary users, both internal and external, to ensure that user perspectives and interpretations are heard, considered, and incorporated appropriately. Evaluation questions should also reflect information that decision makers need. Internally, ELMA M&E staff is committed to engaging with program teams throughout the course of an evaluation, including conceptualizing the evaluation questions, developing a scope of work, providing or reviewing suggestions for evaluators, and reviewing evaluation proposals. During and after the evaluation, staff are actively involved in reviewing progress (including interim findings), and sharing and discussing results. It is important to acknowledge that such a participatory approach—important and often desired—requires considerable time, attention, and collaboration.

Externally, ELMA often engages with a broad range of stakeholders (e.g., government, co-funders, grantees, and stakeholders in the field at large), and involving appropriate key stakeholders in an evaluation is an important precursor to getting strong buy-in and incentivizing the uptake of evaluation results. Stakeholder engagement, however, can be a double-edged sword, with the need to prioritize ELMA's goals in balance with different stakeholder interests, desires, and standards of evidence.

From a grantee's perspective, funder-commissioned evaluations can be particularly fraught, especially given the inherent power imbalance in the funder-grantee relationship. Grantees can feel that evaluations have high stakes, and that future funding decision rests on an evaluation over which they have very little say. It is important that evaluation is understood as an organized decision-making and learning process that involves grantees and their knowledge and perspectives, rather than an exercise that is "done to" them. Clear communication is essential, including being clear about the decisions that the evaluation is intended to inform so that there is no ambiguity about an evaluation's purpose and that it does not cause undue anxiety or create false expectations. In ELMA's *Our Approach to Philanthropy* document, we describe the importance of investing in organizations that are evidence- and data-driven. Mutual buy-in to M&E is important, and very strong resistance to evaluation from the grantee may undermine the evaluation and the investment. It is also important to note that participating in an evaluation can require significant time commitment from grantees, and it is often necessary to resource them to do so.

[4] Gugerty, M. K., & Karlan, D. (2018). The Goldilocks challenge: Right-fit evidence for the social sector. New York: Oxford University Press.

[5] The William & Flora Hewlett Foundation. (2014). *Benchmarks for spending on evaluation.* https://hewlett.org/wp-content/uploads/2016/08/Benchmarks%20for%20Spending%20on%20Evaluation_2014.pdf

5. We are committed to using and sharing M&E results

Creating time and space for internal learning

In the midst of full workloads and busy schedules, it is challenging to intentionally make the time and space to learn from the results from M&E. Opportunities for such reflection—both about the program and the evaluation itself—include brown bags, learning sessions, and evaluation presentations, briefs, or memos. We recommend concluding all evaluations with three questions to elicit reflection: What? So what? Now what? [6] These questions are intended to strengthen the link between evaluation and ELMA's response:

> *What?* entails an understanding of what happened on a grant or cluster of grants, gleaned through the evaluation and complemented with monitoring data, the investment's Measures of Success, and program teams' expertise on the grant.
>
> *So what?* examines why what happened was important, and what lessons—both specific to the intervention and/or context, as well as generalizable for the field at large—can be learned.
>
> *Now what?* looks at future actions, such as what ELMA or grantees will do differently in the future, whether and how grantmaking or program strategies need to change, and what ELMA will do with the evaluation results to influence action.

Tailored communication is a critical to ensure external uptake of results

In addition to informing internal decision-making, evaluations are often commissioned with the goal of having external influence (e.g., crowding in funding, affecting policy or practice). Communicating and disseminating findings strategically are critical to have the desired influence. Knowledge products that emerge from an evaluation cannot be "one report fits all" and must be tailored to the interests and needs of varied audiences, with relevant products (e.g., summaries) made available and accessible for different consumers of the information. In addition, facilitated discussions of the results must take place in a timely manner so that they can be interpreted, understood, and acted on in the context of each stakeholders' needs. As such, each evaluation should be accompanied by a communication plan (see Appendix C) that is developed in partnership with program teams and evaluators.

Every evaluation has its target audience(s); it may not be appropriate or possible to share every evaluation with all possible stakeholders. Communication plans should consider whether it would be appropriate to share findings with the following external stakeholders:

Grantees. Evaluation results have the most immediate implications for grantees' programs. As much as possible, evaluation findings are not only shared with but co-interpreted with grantees.

Partners and other stakeholders. Other stakeholders, such as government partners, funding partners, and/or participants in the evaluation, should receive a relevant summary (and if appropriate, the full report), participate in facilitated discussions of the findings, and be invited to provide input on avenues of communication.

Beneficiaries. Program beneficiaries are often respondents in evaluations but hear nothing back after the evaluation is completed. It may not always be possible or feasible to share the results with beneficiaries, but evaluators should consider the best way to share or give back so that beneficiaries do not experience evaluations as a one-way extraction of information.

---

[6] Human Systems Dynamics Institute. (n.d.). Adaptive Action. https://www.hsdinstitute.org/resources/adaptive-action.html "What? So what? Now what?" broadly correspond to findings (what the evaluation found), conclusions (interpretations of the findings), and recommendations (what actions should be taken as a result of the conclusions), respectively.

Policymakers. Policymakers are a unique audience and have to consider a wide range of interests and issues when adopting or changing policy.[7] They use a broad range of evidence and information from a variety of sources, most of which may not be rigorous scientific studies. When ELMA reviews investments that have the ultimate goal of informing policy or practice by government, ELMA and grantees should have a plan for *using* evidence to advocate for and support government champions on specific policies or practices. While ELMA has built strong relationships with government and policymakers in many countries, using research and evaluation to inform policy is a complex process,[8,9] and no single evaluation will shift policy on its own. It is important to understand what data are especially important for policymakers at which juncture (e.g., cost effectiveness, how to scale a program, acceptability of a program among constituents) and (where possible) include government on discussions around generating evidence, while acknowledging that evidence is a necessary but not sufficient condition for policy change.

The public. ELMA can use multiple platforms to reach broad audiences and the public with evaluation learnings. Our website, conferences, publications, and blog posts are all part of ELMA's communication strategy about key evaluations.

Conclusion

Ultimately, measurement and evaluation serve the goals of generating knowledge, making knowledge usable, and enabling a response to knowledge. These M&E principles provide guidance for how ELMA aspires to approach its M&E practice, while recognizing that learning is in fact a messy and nonlinear process, and that ELMA itself will learn from it and continue to refine its approach.

---

[7] See, for example, Shaxson, L. (2014). Investing in evidence: Lessons from the UK Department for Environment, Food and Rural Affairs. http://www.ksi-indonesia.org/files/1421384737$1$QBTM0U$.pdf

[8] Dhaliwal, I., & Tulloch, C. (2012). From research to policy: Using evidence from impact evaluations to inform development policy. Journal of Development Effectiveness, 4:4, 515-536. DOI: 10.1080/19439342.2012.716857

[9] Jones, H., Jones, N., Shaxson, L., & Walker, D. (2013). Knowledge, policy and power in international development: A practical framework for improving policy. https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/8201.pdf

Appendix A: Evaluation Planning Tool

The Evaluation Planning Tool is a basis for conversations between program and M&E staff on evaluations. It should be used when an evaluation is first contemplated and is a reference for critical questions to answer when planning an evaluation. Being able to answer these questions will help to make sure that evaluations fit our M&E principles, and are sound and well utilized.

FOR EVALUATION PLANNING:	
What decisions need to be made?	
When do they need to be made?	
What specific information is needed to make the decisions?	
What would ELMA do differently if we had the information?	
Who needs the information? (Internal)	
Who else needs the information? (External)	
What/Who are we trying to influence?	
Who will use the results for influence?	
Who will make sure the results are acted on?	

Appendix B: Should We Think About an Evaluation? Checklist

How do we know when we should consider an evaluation? This checklist can help as a tool for decision making. The more check marks an investment has, the more likely M&E and program staff should consider an evaluation.

Importance (the strategic importance of the grant to ELMA)

- ☐ the expected outcome/impact is high
- ☐ large grant investment
- ☐ ELMA's has a long-term commitment to the strategy
- ☐ program's has high potential for scale
- ☐ program's has high potential to improve policy or practice

Influence (the potential for evaluation findings to have influence, especially externally)

- ☐ potential learning from the evaluation is high (e.g., new and/or untested approaches; plausibility of grantee's theory of change; contribution to the evidence base)
- ☐ there is urgency for course corrections or future funding decisions
- ☐ potential for leveraging evaluation findings is high (e.g., providing a basis for securing additional funding)
- ☐ grantee and stakeholder/government concerns are well aligned
- ☐ program's has high potential to improve policy or practice

Appendix C: Evaluation Communication Planning Tool

The Evaluation Communication Planning tool is a reference to help ELMA M&E and communications staff and evaluators to plan for and resource communications around evaluations.

Knowledge Product [Report, summary, presentation, meeting/workshop, etc.]	Audience	Target Date	Lead Contributors	Dissemination Partners or Outlets	Resources Needed