BENCHMARKS FOR SPENDING ON EVALUATION

THE WILLIAM AND FLORA HEWLETT FOUNDATION

Prepared by
Fay Twersky
Amy Arbreton

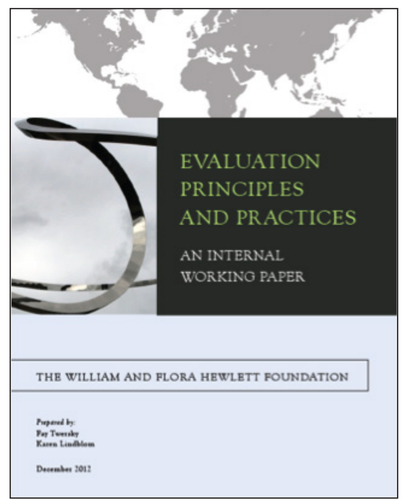
April 2014

TABLE OF CONTENTS

This paper is a distillation from a memo prepared for The William and Flora Hewlett Foundation Board of Trustees. It was presented for discussion and all recommendations were adopted by the board in November, 2013.

1 BACKGROUND

IN NOVEMBER 2012'S BOARD MEETING, A QUESTION WAS RAISED ABOUT HOW MUCH THE Foundation should spend each year on evaluation.[1] The question was prompted in part by a presentation made earlier, at the September Board Retreat, which laid out our plans to incorporate evaluation planning into the programs' strategy development process and to be more systematic about using third-party evaluations. These changes, in turn, call upon us to articulate benchmarks for the expected costs of evaluating our grantmaking programs' progress and outcomes.



Under Paul Brest's leadership, the Foundation developed a rigorous, nine-step process for designing grantmaking strategies that we call "outcome-focused grantmaking" (OFG).[2] While OFG recognizes the importance of evaluation, the practice remained loosely defined and inconsistently employed until 2012, when the Foundation developed the "evaluation principles and practices" discussed at the September 2012 retreat.

In order usefully to inform ongoing decision making in areas as diverse as performing arts, education, environment, and global development, these principles and practices must be pragmatic and flexible. To support their implementation and ensure consistent, high-quality evaluation across programs, we hired the Foundation's first evaluation officer in 2013. Programs remain responsible for commissioning their own evaluations, with the evaluation officer providing technical assistance in design, planning, and analysis.

By providing information about what is or is not working, evaluations can improve not only our funding decisions but also the work of our grantees, with whom we share the results and who likewise learn from them. Beyond even this, by providing evidence of success or failure in areas where others also work, evaluations can help to improve practice generally. Who benefits from an evaluation affects how we pay for it. If an evaluation benefits others, whether this be particular grantees or a field generally, it is recognized as a "direct charitable activity" by the IRS and can be budgeted out of grant dollars. If not, it is considered an administrative expense and will be funded from our administrative budget.

In practice, the line between work that has use outside the Foundation and work that is wholly internal is often fuzzy (which means we have some discretion about funding an evaluation from our grant budget or our administrative budget). For present purposes, our focus is on evaluations that at minimum influence and improve our own funding decisions, recognizing that these are sometimes useful to outside actors and regardless of their funding source.

1 For the purpose of this memo, "evaluation" refers to studies and assessments conducted by a third party.

2 http://www.hewlett.org/uploads/documents/Outcome_Focused_Grantmaking_March_2012.pdf

2 TRENDS IN EVALUATION SPENDING

OVER THE LAST THREE YEARS, WE HAVE SPENT BETWEEN 0.7 PERCENT AND 1.2 PERCENT OF programmatic dollars on evaluation. If we exclude our grant to ClimateWorks from the calculation (to avoid the distortion created by its unusual size), the proportion becomes slightly higher, placing our evaluation spending between 1.0 percent and 1.7 percent of programmatic dollars.

The question is, how much should we be spending?

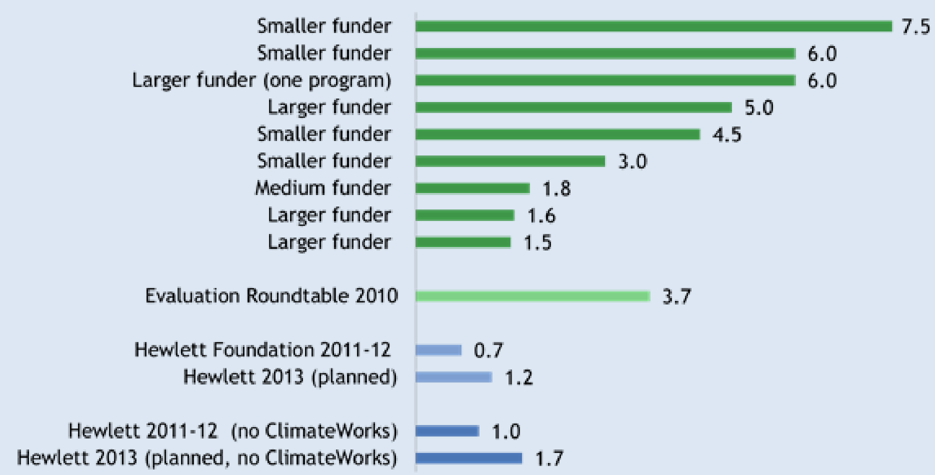
Conventional wisdom long held that a serious commitment to evaluation required spending on the order of 5 to 10 percent of programmatic budgets. In 2010, the Evaluation Roundtable (an association of evaluation professionals) carried out a benchmarking study and found the actual foundation industry norm to be 3.7 percent. Spending varied with, among other things, the size of the foundation. Because the costs of evaluation do not rise proportionately with program costs,[3] larger foundations typically spent a somewhat smaller proportion of their budgets on evaluation than did smaller foundations, even if they were equally serious about evaluation.

The Evaluation Roundtable has not been able to repeat its benchmarking, so in preparing this memo, we conducted a benchmarking effort of our own by polling other foundations that we believe have strong evaluation practices.[4] We asked how much they typically spend on evaluations intended to inform their own grantmaking and strategic decisions. We asked them to exclude research studies like randomized controlled trials that are foremost intended to inform a field. Figure 1 shows our proportional spending compared with these institutions.

In examining these figures, bear in mind that most foundations (including ours) do not have good systems for tracking evaluation spending and that we all classify and calculate evaluation costs in slightly different ways. Hence, colleagues at other foundations prefaced their replies to our inquiry with an apology that the data were not as rigorously compiled as they would have liked, with some suggesting that their reports were closer to an estimate than an actual accounting. Even with that caveat, however, the

3 There is a basic threshold cost for a decent evaluation. Expenses associated with research and instrument design, site visits, data analysis, report preparation, and project management don't rise proportionally with program expenditures.

4 We drew largely upon an active group of foundation evaluation professionals who agreed to share their evaluation spending with us. For the purpose of this public version of the memo, we have "blinded" the names of the other foundations.



**Other foundations' data were reported in 2013. Time periods covered by the data may vary. "Larger funders" are those with more than $200M in grants in 2011. "Medium funders" made $100-200M in grants, while smaller funders made less than $100M.*

FIGURE 1 Percent of budget spent on evaluation relative to overall grant budget*

information is roughly accurate, and it puts the Hewlett Foundation at the very low end of the spending curve.[5]

A number of factors explain this result, including

1. Size. We are a large foundation, and our spending resembles the proportionately lower spending of other big foundations.

2. Focus. We have generally focused on getting our initial strategies right, thus concentrating our energy and efforts at the front end of the OFG process.

3. Time. We are leanly staffed, and program officers have limited time to commission and oversee evaluation work.

4. Evaluation know-how. Program officers are hired for their domain expertise, and few come with evaluation experience or skills. Unlike the foundations we consulted, we have not had technical evaluation staff (until this year). This lack of evaluation know-how, in turn, may have inhibited staff from commissioning evaluation efforts in the first place.

5. Pressure on administrative budget. Desire to keep administrative costs low, given factors two through four above, may have encouraged underinvesting in evaluation.

5 There are instances in which we are able to take advantage of evaluation efforts funded by others. The recent evaluation of the Think Tank Initiative, which was paid for by the Gates Foundation, is an example. Just as often, however, other funders free ride on our efforts. In terms of overall spending, we believe these costs and savings are probably a wash.

3 EVALUATION AS LEVERAGE

WHICH RAISES THE IMPORTANT QUESTION: WHY SPEND ANYTHING ON EVALUATION? WHY NOT keep administrative costs lower or use these resources for grants? What benefit do we get from doing this work?

The answer, in a word, is leverage. We invest a little to learn a lot, and in learning we make our grant dollars more effective and more efficient. We gain information that facilitates superior grant allocations going forward, helps us adapt current grantmaking, reveals promising new directions for work, and so on. Evaluation helps us make our grant dollars go farther and do more, and it does so in a way that we believe makes the net benefits well worth the expenditure.

Examples of evaluations that provided important, leveraged learning can be found across all of our programs. Consider some very recent illustrations.[6]

A. Open Educational Resources

Open educational resources, or OER, comprise teaching, learning, and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use by others. Since 2009, the Education Program has funded four national grantees to increase awareness of OER among policymakers and to promote government procurement policies and implementation guidelines that are favorable to OER. The Foundation also invested to create a coalition of like-minded nonprofits working on this issue.

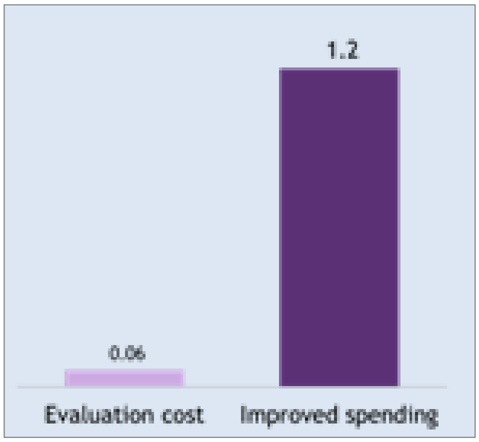


FIGURE 2 OER evaluation leverage ($M)

In 2013, after the OER movement achieved several significant policy victories at both the federal and state levels, the Education Program commissioned a $63,000 evaluation to assess the $1.2 million it had spent over the previous four years—seeking to understand the role our grantees had played in securing these victories, the effectiveness of the coalition structure, and the best approach to future investments. The evaluation found that the policy victories were produced chiefly by the efforts of "inside champions" in the policymaking bodies. While the grantees' efforts were focused on broader advocacy, this was having less of an effect. The

6 Leverage is calculated for each case by comparing the cost of the evaluation with the grantmaking dollars affected by the evaluation findings. In some cases, an evaluation might inform only future grantmaking in an area. In other cases, an evaluation might inform current grants as well as future grantmaking.

evaluation also concluded that the coalition was not an efficient structure to produce the desired outcomes.

Based on this feedback, the Education Program has shifted its approach for its next $1.2 million investment. The program will focus on developing inside champions directly, while tying off some of the support for the less effective part of the strategy and rethinking the coalition structure. As Program Officer Vic Vuchic explains, the evaluation provided “not just capital efficiency, but also strategic discovery.”

B. Western Conservation

During the past five years, the Western Conservation team spent approximately $60,000 on evaluation each year, continually using what it learned to refine its work in real time. For example, a 2011 evaluation of grantmaking to lessen the impact of fossil energy development in the West concluded that significant new investments were needed in communications capacity. This led to grants to two new organizations that proved highly successful in supporting our efforts.

The Western Conservation team’s mid-course strategy evaluation provides a still clearer illustration of the kind of leverage that can be produced by a high-quality evaluation. This $200,000 effort assessed the effectiveness of the team’s approach from the perspectives of both science and policy. Results have led to important adjustments that will significantly improve our conservation strategy, which will spend approximately $21 million per year in its next five-year phase.[7] Given the Boreal co-investment agreement the team expects to strike with two other foundations as a result of this evaluation, the grant dollars affected will be more on the order of $27 million per year.

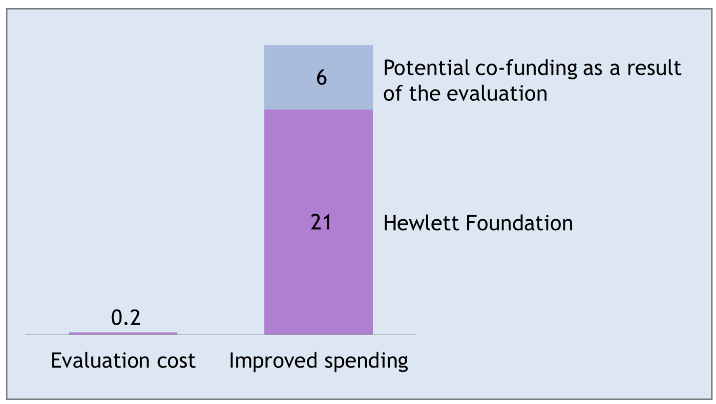


FIGURE 3 Western conservation evaluation leverage ($M)

C. Community Leadership Project

The Community Leadership Project was launched in April 2009 as a partnership between the Packard, Irvine, and Hewlett foundations. Its purpose is to build the leadership capacity of small and midsize nonprofits serving low-income people and communities of color in three California regions. In Phase I, the three foundations made grants totaling $10 million to twenty-seven intermediary organizations, which re-granted funds and provided assistance to a diverse set of community organizations.

The Project brought with it a number of challenges—including complex dynamics between funders, technical assistance providers, and grantees; the

7 One might imagine the leverage being five times the amount listed here since the $27 million is an annual amount. Thus our leverage estimate in this instance is on the conservative side.

need for cultural sensitivity in dealing with diverse populations and organizations; and grantees' limited staff time to participate in Project activities. An evaluation was commissioned early on to inform the Project's uncertain theory of change and figure out ways to improve its implementation.

The $500,000 evaluation revealed many insights about how to provide effective, culturally competent assistance to these organizations and their leaders. The lessons learned—both what to do and what not to do—have significantly improved the Project's execution. These lessons include refinements in the criteria used to determine a potential grantee's readiness to participate and improved ways to anticipate and handle logistical challenges that small nonprofits uniquely face when asked to attend meetings or otherwise participate in the initiative. The evaluation also provided critical feedback that led to a more focused design for Phase II, which provided a second $10 million in grants.

As a result of this evaluation, Phase II involves far fewer intermediaries; includes highly integrated technical assistance along with general operating support; focuses more on smaller organizations; and gives greater attention to financial sustainability and leadership development.

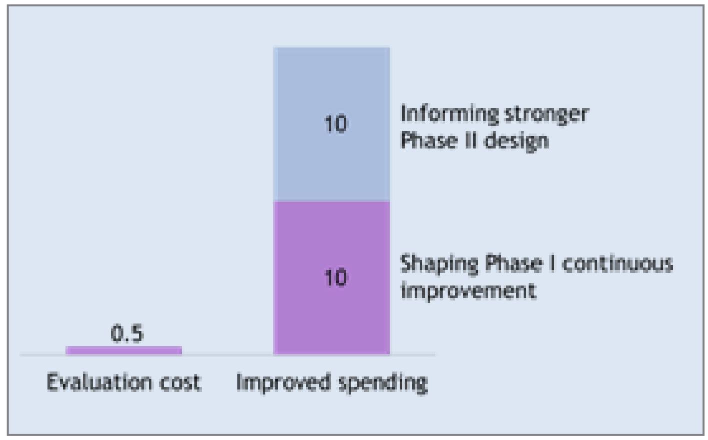


FIGURE 4 Community Leadership Project evaluation leverage ($M)

The math is simple: if an evaluation whose cost is equal to 5 percent of the grants under review delivers results that are 10 percent better as a result of the evaluative information, it is a worthwhile expenditure; if it delivers a 50 percent improvement, the return is extraordinary. If it frees up funds not otherwise being spent well, it allows for new experimentation and learning. In other words, a small amount of evaluative information can result in directional guidance that generates better outcomes through smarter spending.

4 ROOM FOR IMPROVEMENT

WHILE THESE EXAMPLES (AND MANY OTHERS WE COULD SIMILARLY OFFER) ILLUSTRATE THE Hewlett Foundation's effective use of evaluations, we can do still better. In particular, we believe we can capture additional value from evaluation if we (1) focus more serious attention on high-stakes work, (2) integrate evaluation concerns into the design of our strategies, and (3) strengthen our methodology. We discuss these briefly below.

A. Focusing on High-Stakes Work

ClimateWorks 1.0 was the Hewlett Foundation's single largest grant ever in the amount of $500 million over 5 years. Yet rather than commission an independent evaluation of the initiative, we deferred to ClimateWorks to "own" the evaluation. ClimateWorks spent $400,000 (.08% of our grant and .04% of the total $1.1 billion in grants provided by three foundations) on an evaluation effort designed to track the network's success in reducing carbon emissions—far too modest an investment and narrow a scope of questions for an initiative of this size. The resulting evaluation fell short of providing the kind of timely, critical information needed for ongoing learning and course correction. For instance, the evaluations did not consider questions of organizational health and development—an almost certain stumbling block when launching a big intermediary start-up—and despite the regular evaluation reports, there was minimal application of analysis into action.

We believe that an independent and more robust evaluation of ClimateWorks, one in which the Hewlett Foundation and the other funders held at least part of the contract, could have surfaced challenges sooner—perhaps in time for an earlier course correction. Spending a little more early on would have been a wise investment, possibly saving a lot of time (and money) in keeping this climate initiative on track.

Good news: we are planning for stronger evaluation of the Climate Change Initiative going forward, involving more funders and in close coordination with the ClimateWorks 2.0 team. The evaluation will draw on a wider base of methods for data collection and be designed to answer key questions in a pragmatic and rigorous fashion.

B. Planning Early for Evaluation

A key principle we adopted this past year is to incorporate evaluation into the process of strategy development. This does not mean that we evaluate everything. It simply means that early in the process of developing a strategy we consider the kinds of information we'll need to test our assumptions and to determine success. This allows us to plan in a stepwise fashion for data collection and use.

To understand the importance of early evaluation planning, consider an example in which we failed to do so: the Philanthropy Program's Nonprofit Marketplace Initiative. By the time we commissioned a first evaluation, eight years into the strategy, we had lost the ability to collect good baseline data. In addition, an inadequate focus on evaluation at the outset had contributed to lack of clarity about what would constitute success. When we did evaluate, the results were disappointing.[8] Had the Foundation considered evaluation from the beginning and collected actionable data along the way, we might have positioned ourselves to address key challenges or change course earlier.

Good news: as we develop new strategies, program teams now routinely plan for evaluation from the beginning. Examples include the strategy refreshes in Western Conservation and International Reproductive Health and Family Planning.

C. Strengthening Methods

Finally, many of our past evaluations could have been stronger methodologically, which would have added greater certainty and nuance to the findings. The Foundation has commissioned evaluations whose guiding questions were not crisply articulated or that relied too heavily on qualitative point-in-time interviews as their sole method. Much evaluation work, especially in the policy and advocacy arenas, must necessarily rely on qualitative information; there are, however, ways to strengthen these evaluations by using appropriate metrics or multiple methods over a series of points in time.

Good news: current evaluations are more commonly incorporating multiple methods into their original designs. We expect our new evaluation officer to be especially helpful in this regard.

8 http://www.hewlett.org/library/call-proposals/evaluation-nonprofit-marketplace-initiative

5 RECOMMENDATIONS FOR SPENDING TARGETS

TO ENSURE WE LEARN WHAT IS NEEDED TO MAKE TIMELY, INFORMED, SMART DECISIONS—AND in this way to get the greatest leverage from our grant dollars—we recommend a modest increase in the amount we currently spend on evaluation. We need not aim for the top of the spending chart in Figure 1: spending only what is needed to make the adjustments described above will significantly improve the quality of our evaluation practice and so make the rest of our grantmaking that much better.

There is no mathematical formula or simple process to use to establish a budget target, so we approached the task in two ways. First, we looked at current spending and the quality of work and results it has purchased and made an educated guess. As noted above, our 2013 evaluation spending looks to be 1.2 percent of our grant budget (including ClimateWorks). Realistically correcting for the sorts of improvements discussed above suggests increasing our spending to approximately 2 percent. Depending on how much value the evaluations add, we could increase our spending incrementally from there.

As a check on this top-down estimate, we conducted a corresponding bottom-up budgeting exercise with colleagues from the Redstone Strategy Group. In this, we sought to operationalize the evaluation principles and practices laid out in our 2012 working paper. According to this exercise, we could usefully increase our spending to 2.3 percent of program spending.

Putting this all together leads us to make the following recommendations for the Board's consideration:

1. Over the next three years, the Foundation should aim to increase its spending on evaluation to approximately 2 percent of program spending. (In actual dollars, this means increasing our spending from $4.0 million in 2013 to $6.6 million in 2016.)

2. The focus of our increased spending should be on improving the quality and practicality of our evaluations (as opposed to simply funding more of them), thereby producing insights that add value and lead to better grantmaking.

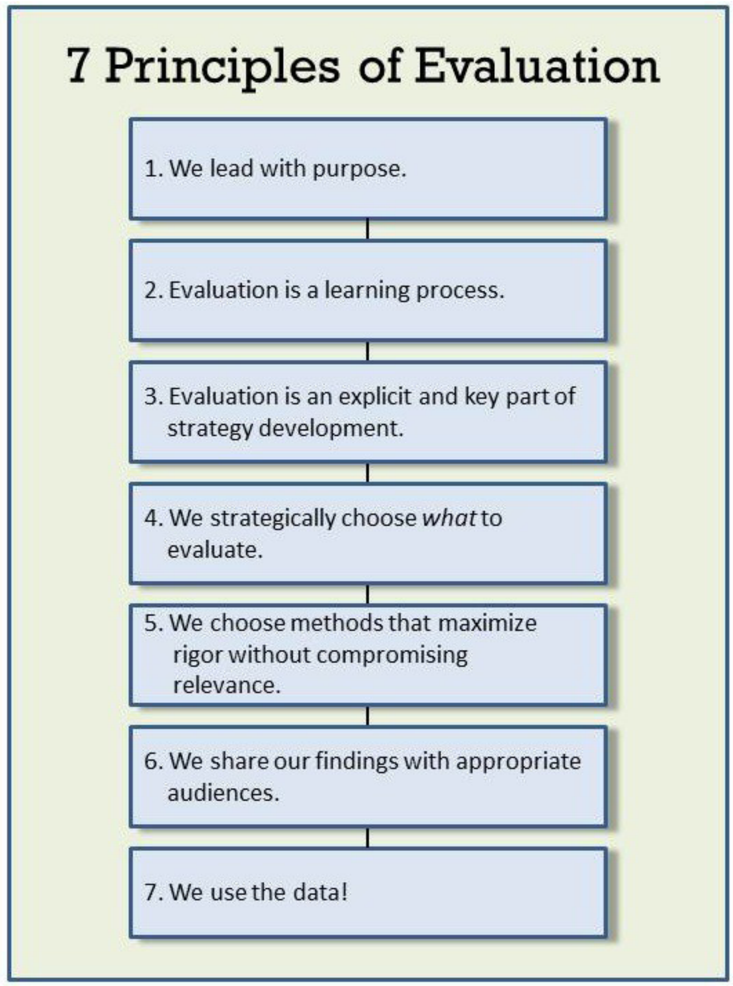
3. We will continue to pay for evaluations with a mix of administrative and grant budget funding. This means the additional spending should have minimal impact on our administrative costs, as both grant-funded evaluations and contracts that qualify in whole or in part as direct charitable activities are treated as coming from the grants budget without affecting administrative overhead.

4. We should improve our systems for tracking evaluation expenditures so we have more accurate data on overall costs and on the costs associated with different types of evaluations.

5. We will assess the value we are deriving from evaluation and report back to the Board in three years.

APPENDIX 1 THE WILLIAM AND FLORA HEWLETT FOUNDATION'S EVALUATION PRINCIPLES

We aspire to have the following principles guide our evaluation practice:



1. We lead with purpose. We design evaluations with actions and decisions in mind. We ask, "How and when will we use the information that comes from this evaluation?" By anticipating our information needs, we are more likely to design and commission evaluations that will be useful and used. It is all too common in the sector for evaluations to be commissioned without a clear purpose and then to be shelved without generating useful insights. We do not want to fall into that trap.

2. Evaluation is fundamentally a learning process. As we engage in evaluation planning, implementation, and use of results, we actively learn and adapt. Evaluative thinking and planning inform strategy development and target setting. They help clarify the evidence and assumptions that undergird our approach. As we implement our strategies, we use evaluation as a key vehicle for learning, bringing new insights to our work and the work of others.

3. We treat evaluation as an explicit and key part of strategy development. Building evaluative thinking into our strategy development process does two things: (1) it helps articulate the key assumptions and logical (or illogical) connections in a theory of change; and (2) it establishes a starting point for evaluation questions and a proposal for answering them in a practical, meaningful sequence, with actions and decisions in mind.

4. We cannot evaluate everything, so we choose strategically. Several criteria guide decisions about where to put our evaluation dollars, including the opportunity for learning; any urgency to make course corrections or future funding decisions; the potential for strategic or reputational risk; size of investment as a proxy for importance; and the expectation of a positive expected return from the dollars invested in an evaluation.

5. We choose methods of measurement that allow us to maximize rigor without compromising relevance. We seek to match methods to questions and do not routinely choose one approach or privilege one method over others. We seek to use multiple methods and data sources when possible in order to strengthen our evaluation design and reduce bias. All evaluations clearly articulate methods used and their limitations.

6. We share our intentions to evaluate, and our findings, with appropriate audiences. As we plan evaluations, we consider and identify audiences for the findings. We communicate early with our grantees and co-funders about our intention to evaluate and involve them as appropriate in issues of design and interpretation. We presumptively share the results of our evaluations so that others may learn from our successes and failures. We will make principled exceptions on a case-by-case basis, with care given to issues of confidentiality and support for an organization's improvement.

7. We use the data! We take time to reflect on the results, generate implications for policy or practice, and adapt as appropriate. We recognize the value in combining the insights from evaluation results with the wisdom from our own experiences. We support our grantees to do the same.