grantcraft

PRACTICAL WISDOM FOR GRANTMAKERS

Making Measures Work for You

Outcomes and Evaluation

Evaluation Techniques: A Series of Brief Guides

This guide was written by Craig McGarvey. It is part of the GrantCraft series.

Publications and videos in this series are not meant to give instructions or prescribe solutions; rather, they are intended to spark ideas, stimulate discussion, and suggest possibilities. Comments about this guide or other GrantCraft materials may be sent to Jan Jaffe, project leader, at j.jaffe@grantcraft.org.

To order copies or download .pdf versions of our publications, please visit www.grantcraft.org.

Contents

What are "outcomes," and what are the implications of trying to measure them?

Outcomes are the observable results of programs that are created and funded in hopes of making a difference in the world. The case for measuring outcomes is hard to refute: if you can't measure it, say the proponents of evaluation metrics, how can you know it's happening? An effective organization or project, they argue, has measurable effects. The steps you follow to make change — your *outputs* — are important, but the results of the process — the *outcomes* — are what matter in the end.

To illustrate, here are a few concrete examples of outputs and some outcomes to which they are related. A full-scale evaluation would likely ask questions about several interrelated outcomes.

- The leaders of a nonprofit organization want to help a population of chronically underemployed people find decent jobs with good prospects. They research the industry sectors that are hiring and build relationships with human relations managers. They develop and implement a training course that includes both hard skills (how to run the equipment) and soft ones (how to interview, how to get to work on time). The training classes are full, and an evaluation demonstrates that the operation is run creatively and efficiently. But are the graduates of the program getting jobs? What if they get first jobs but are unemployed again a year later? The organization would want to know those things in order to improve its work.
- A youth program seeks to reduce teenage smoking locally. The program has engaged a top-notch ad agency that is designing clever public service announcements, and the local television station is running the ads, all pro bono. The organization has developed a superb curriculum, using a highly regarded student peer counseling approach, and the high school has adopted it. The program is reaching its target population, but is the incidence of teenage smoking going down?
- A foundation plans to achieve systemic impact on agriculture by supporting sustainable farming practices and encouraging local ownership and consumption, while also helping producers stay competitive globally. The strategies are multifaceted and complex, including advocacy, education, communications, and labor and community organizing. Ultimately, the funder wants to know what effect its work is having on policies and practices. For instance, how many acres of farmland have been converted to sustainable stewardship?

The point is to put a spotlight on what a program is actually accomplishing — what difference it's making — rather than simply looking at the interventions themselves. But how do we actually *measure* those outcomes? And, in a complex world, how do we confidently draw connections between the inputs our programs have contributed and the outcomes we ultimately observe?

Answering those questions reveals the power of the approach, and also the rub. In the past few years, evaluation metrics have inspired lively debate in philanthropy and the nonprofit sector — a debate that touches on the nature of knowledge building, the relationship between accountability and learning, the difference between the change itself and the proxy being used to measure the change, and, indeed, the very purpose of the charitable arena.

Where the examples in this guide come from

For reflections on outcomes measures and how they are being used, we talked with grant makers, evaluators, and consultants involved with projects supported by many foundations, including the following:

- A national foundation that sees evaluation and knowledge development as part of its strategy to build the organizational capacity of a select group of grantee organizations
- A regional foundation that uses evaluation and outcomes identification to help organizations develop a "culture of reflective practice"
- A family foundation that asks grantees to name the outcomes they aspire to, even if measuring them ultimately proves elusive
- A corporate foundation with international scope that gives priority to proposals that "clearly articulate the planned impact of their efforts, as well as the metrics by which that impact would be measured"

This guide is not intended to pursue the debate at its most rarefied levels, where professional researchers argue fine shades of meaning. Rather, our objective is to distill core questions, define terms, and help grant makers gain real insights into the effectiveness of the work they fund. As one grant maker explained, you don't need to be an "accountability junkie" to work confidently with evaluators and grantees in the quest for answers about outcomes.

To set our feet on the solid ground of practice, the guide begins with the story of an organization that was asked by its funder to conduct an outcomes-oriented evaluation. The next section reviews some common questions that come up repeatedly as grant makers think through the use of outcomes measurement with grantees, evaluators, and colleagues. The last section sketches out seven tensions in the debate about outcomes, and how some grant makers have negotiated them.

To learn more . . .

The following sources offer examples of how foundations and nonprofits are using outcomes measurement.

Articles and chapters

- John Sawhill and David Williamson. "Measuring What Matters in Nonprofits." *McKinsey Quarterly* 2001 (no. 2). (available at www.mckinseyquarterly.com)
- Melinda Tuan. "Cultivating a Culture of Measurement," chapter 5 in *Funding Effectiveness*. Jossey-Bass, in association with Grantmakers for Effective Organizations, 2004.

Online tools

- B2PCommerce (www.b2p.com). B2P's *Impact Manager* software tool is designed to track portfolio performance, developed in collaboration with Community Foundations of America.
- REDF (www.redf.org). The *Social Return on Investment Collection*, includes reports, cases, and updates on SROI. See also the REDF paper, "An Information OASIS," which calls on funders to get involved in measuring social outcomes.
- Women's Funding Network (www.wfnet.org). The network's *Making the Case* utility, enables organizations to measure and document social change over time. (Available only to members.)

Books

- Marc T. Braverman, Norman A. Constantine, and Jana Kay Slater. *Foundations and Evaluation: Contexts and Practices for Effective Philanthropy*. Jossey-Bass, 2004.
- Michael Quinn Patton. *Utilization-Focused Evaluation: The New Century Text*. Sage Publications, 1997.

Getting Past the Evaluation Jargon

As grant makers, we want evaluation and assessment techniques that help document and analyze the work we support in ways that are meaningful to our foundations, grantees, and wider field or community. To help grant makers weigh the advantages of different approaches, GrantCraft offers **Evaluation Techniques: A Series of Brief Guides**. Each guide explains the basics of one technique, answers common questions about its use, describes how some grant makers are applying it, and includes a list of resources for readers who want to learn more. See **www.grantcraft.org** for other titles in the series.

A mini-case study: Using outcomes evaluation to improve learning

When her program officer raised the prospect of doing a formal evaluation of her charter school, the executive director rolled her eyes. Everybody wanted to evaluate her students, and she was tired of it. The curriculum was based on developmental principles and built around authentic experience: students created projects and took them to completion in performance — by making a presentation to the city council, for example, or by convincing the mayor to accept their researched recommendation. With the help of a post-doctoral student, the school had earlier articulated a clear theory of their "action learning" approach.

The school already thought of itself as a learning organization. A formal evaluation seemed excessive; the senior staff who worked at the school were not certain what they could learn from it. Their foundation program officer, though, believed thoroughly in evaluation as a "transparent approach to uncovering why folks practice in the way they do." The evaluation would be a "generative process," a "tool to investigate how to retool" and "prepare for the next turn," looking to a sustainable future for their organization.

The evaluation involved identifying outcomes and indicators of the outcomes, then trying to see if the results were being achieved. "We started with 52 variables!" recalled the director (that number was eventually winnowed down). The work of outcomes identification was challenging but rewarding. Staff members had "great conversations about what outcomes look like. Also, we asked why we believe that this kind of professional behavior will lead to these outcomes. It was an excellent learning experience."

Some examples: "We claimed the kids had experiences," explained the director, "but we were essentially providing inputs. Did the kid actually have the experience? Could we track it? Would the kids report it?" Asking students to accept risk was one of the school's "5 Rs," the director went on, but "what behaviors would provide evidence?" The overall goal of the school was to "create effective citizens, who know what they believe, think and feel, and can act on behalf of themselves and others," but what was the school actually doing in each of those three realms?

In addition to substantial ongoing core support, the school spent two years and approximately $300,000 to evaluate its entire program. Evaluation questions were divided into outputs ("Are students getting the experiences we intend to be the core of our program?") and outcomes ("Are those experiences producing results in cognitive, social, civic/moral, and emotional areas?"). They looked for community outcomes, too, among the parent body. Data were collected through observation, interviews, group reflection, and survey reporting.

There were many challenges along the way:

- The evaluation was "counting on students to be good at self-reporting," noted the director, but that did not prove to be particularly true: "There were times when we could see that the kids were growing, but they didn't see it." "Separate from other sources of data," student self-reporting "did not stand alone. It told us something about their experience, but not enough about outcomes."
- "Making meaning of the data" from student surveys "required a heck of a lot of contextualizing." For example, at one point the surveys produced a statistically significant result that students felt they were not getting "opportunities to be leaders in social action projects." It didn't make any sense to the staff. Digging back through the surveys, they found that the forms had been filled out the day after a project had been canceled because of something outside the school's control.
- Designing the surveys was methodologically challenging: "How do you describe things so young people will understand?" After the first year, it was clear that students had not adopted the language of the evaluation, so the adults started using it explicitly in daily work. The surveys had to be shortened; "after a couple of pages, students were just circling things."
- "Another reality," reported senior staff members, was that, although the school tried hard to get students and coaches to "take ownership" of the evaluation, "it never became part of life — it was just something that was passed out and collected." The exceptions were certain instruments, such as youth feedback forms and personal skills reports, which were "useful to staff in doing their regular jobs."

So, was the experience a failure? Not at all. Reflecting back on the evaluation, staff members said that it inspired them to work harder to "embed assessment and learning" into the school's culture. The step of "identifying measures of success" has been adopted "very organically." A big result was a renewed focus on adult learning at the school, and more intentional links between adult and student learning.

From the foundation program officer's point of view, the evaluation achieved success: "Earlier, when they talked about the school, you would hear about strengthening academic skills." In the course of the evaluation, school staff began to shift in their language, saying they hoped to give young people the skills to be "intellectually agile, so they can find information and be good problem solvers." Among adults, the program officer said, the school took on a "culture of evaluation. … The evaluation helped them become more reflective practitioners."

Common questions about outcomes measurement

Here are some of the questions and answers we heard when speaking with grant makers and grant seekers about outcomes.

Is the debate over measurable outcomes just the latest incarnation of questions about quantitative versus qualitative evaluation?

In some ways, yes, although some people think the issue is a red herring. Says a foundation president, "Most important to me, as a funder, is that the organizations in which we invest ask themselves — often — how they are doing, and how they know it, and reflect on that information to adjust and improve their program design and delivery. Are numbers part of it? Certainly. But so is client feedback, environmental and economic factors, and so on."

Others agree that reflection is what counts. A grant maker whose foundation stresses outcomes explained: "You have to be interested in outcomes, otherwise you're only looking at process. Setting forth expected results ahead of time, developing indicators to assess success — those things force programs to say what success looks like. But the value is in the doing of it."

The real issue seems to be balance: balance among various outcome measures, both loose and hard, and balance between qualitative, narrative descriptions on one hand and quantitative on the other. Ideally, they should complement one another. "No numbers without stories; no stories without numbers," said one evaluator.

What do I say to someone who truly believes his or her work can't be reduced to concrete measurement?

Often the exercise is less about reducing things to numbers than, in the words of a grant maker, "asking yourself, what information do I need in order to know I am doing a good job?" "Let's not even call it evaluation!" he insisted. "At the end of every day, I come home with some intuitive sense of how I've done. How can I figure out how I've come to this hunch?"

That sentiment was echoed by a senior program officer at a foundation that invests heavily in outcomes evaluation. His standard speech to grantees includes a simple explanation: "You know you're effective. That's what gets you out of bed in the morning. But I need more than a gut feeling. I need four or five things that will tell our board that you're kicking butt."

Measuring Intangibles

It's not uncommon for a program to aim to change such difficult-to-measure outcomes as peoples' thinking, attitudes, or feelings. In school reform, for example, a large body of evidence points to a connection between student performance and teacher expectations. A related body of evidence points to a connection between teacher expectations and teachers' own feelings of being supported as professionals within their schools. It's reasonable, then, to assume that supporting teachers can have a positive effect on student performance.

The planners of one foundation-funded initiative drew on that thinking to design a program to create a large number of small high schools, with the long-term objective of improving students' graduation rates. The small schools have received some very specific help with building strong, cohesive professional communities. But have those communities actually taken shape? And, if so, how can one measure whether teachers feel supported within them?

For answers, the program built an annual teacher survey into a larger evaluation design. As part of that survey, teachers were asked about professional respect and support from many different angles. Three of the survey's many items asked them to rate their level of agreement with the following statements:

- Teachers in this school are evaluated fairly.
- Necessary instructional materials are available as needed.
- Staff members are recognized for a job well done.

As one of the evaluators explained, the responses don't necessarily say a lot about a particular small school, "unless you get a very high response rate." Still, she said, the data can be an important part of a bigger picture. The teacher survey offers "a bit of evidence that makes sense with all kinds of qualifiers around it," such as qualitative reports by evaluators and program staff, surveys completed by principals and students, and student performance data.

Being able to tell it is the goal. "To anyone trying to make change," said a foundation vice president, "it is terribly helpful to be able to articulate why they are doing it this way, what they are accomplishing, and how. It is talking about not only doing right things but doing things right, being clear about what you are aiming for. It's testing your compass."

Can you always know up front what specific outcomes you should be aiming for?

You can't, but that doesn't mean you can't get started with your best analysis. The danger comes from sticking too stubbornly or too narrowly to originally identified outcomes. "We understand that if in the middle of the evaluation you find it's better to measure something else," said a funder, "you should change it. Don't get hung up."

"It's a continuous process of feedback you are looking for," explained an experienced grant maker. "You may end up somewhere you hadn't planned, but that's okay. You want timely and useful data so you can make adaptive changes."

Another grant maker emphasized a related point: "No matter what measures we pick, they're always shaped by what we believe is most important. It's useful to ask ourselves about those beliefs — and also to ask around, find out what others think, and stay open to other points of view about what is a meaningful outcome measure."

Where has the outcomes measurement push come from? Is it, in the end, about transparency and accountability?

According to most grant makers interviewed for this guide, the recent impetus has come from two sources. The new wealth from the economic boom of the late nineties created new foundations, many of them led by business people who were accustomed to the use of metrics and interested in demonstrating effectiveness in their grant making. And, yes, concerns about philanthropic accountability — and a desire among foundations to get a grip on their own affairs before government steps in to impose further regulation on the sector — have added momentum. "The accountability wave is coming," said one foundation official, "and it will be like stepping on a garden rake — whappo!" In his view, foundations need to "crack the nut" of demonstrating their own effectiveness, and the sooner the better.

If some of this is being driven by the accountability surge, are foundations running the risk of putting an unacceptable degree of pressure on their nonprofit grantees?

Although this particular worry accompanies any attempt at evaluation, it may be especially relevant to outcomes evaluation. In the words of a grant maker committed to this approach, "You're going to get good news and bad news." The grantee is being asked to be transparently honest in the presence of a funder, and that's a vulnerable position to occupy.

Even when funders are at their most sensitive, the challenge can be great. One grant maker who worked to develop data-based assessment systems with her grantees writes about the "slippery slope" into a "punitive system." "The slightly unintended effect of having great social outcome measurement systems is that the numbers they are reporting are not quite as positive as what funders are used to seeing," she noted. Just as "embracing a culture of measurement" is a new way of doing business for grantees, foundations, too, will need to change practice, in how they "analyze data, reflect on data, and react to the data for future funding decisions."

Almost everyone interviewed for this guide felt that the risks of honest interaction are well worth taking. "The ones with the metrics will get the money," said a funder, "but the nonprofits have to literally own the information."

It comes down to establishing an authentic partnership, in which the data facilitates dialogue. One funder summarized it this way: "You want to learn together? The important thing is relationship, relationship, relationship."

What is the role of technology in outcomes evaluation?

Several foundations and for-profit ventures are enthusiastically pursuing the promise of technological tools for tracking and reporting on the performance of nonprofit grantees and overall foundation portfolios. For example, B2P's Impact Manager software, available through the Community Foundations of America, helps funders and grantees articulate outcomes and set and track measures. Other funder-generated evaluation tools include Making the Case, developed by the Women's Funding Network, and OASIS (Ongoing Assessment of Social ImpactS), created by REDF (formerly known as The Roberts Enterprise Development Fund).(Further information and links to relevant websites may be found on page 3 of this guide).

What exactly is social return on investment, or SROI?

Social return on investment (SROI) describes the contributions nonprofits make to social value in communities: a double, or, more accurately, multiple bottom line. Several foundations have been working to come up with reliable, quantitative methods for analyzing doing good. For example, in its work with nonprofit organizations that run small businesses with social missions, REDF has developed a framework for measuring SROI, using six metrics: enterprise value, social purpose value, blended value, enterprise

> "No matter what measures we pick, they're always shaped by what we believe is most important. It's useful to ask ourselves about those beliefs — and also to ask around, find out what others think."

index of return, social purpose index of return, and blended index of return.

When is an outcomes evaluation approach most appropriate?

Outcomes evaluation is almost always appropriate, say many funders, although they would tend to use it more flexibly with a start-up or exploratory effort than with a proven model. It can be especially valuable when a mature organization is trying to refine its work, replicate its operations, or test a specific intervention.

One major philanthropy has invested in building the capacity of a field as a whole by making grants to strengthen evaluation at several organizations. The foundation saw the field as "weak and underperforming" but was confident it could be "built into a critical mass of high-performing organizations" by increasing their ability to do outcomes-oriented evaluations. "We were dropping pebbles into a marsh," explained a foundation official, "but we felt a footpath could be created."

There is also an issue of proportionality. The executive director of a management consulting organization warns against imbalance at both ends of the spectrum, whether trying to get "results in public education with $500,000" or "spending $5 million on a multi-year evaluation with a prestigious research firm" that couldn't answer whether the initiative was effective. The watchword is balance.

How much should it cost, and do I always need an outside consultant?

"You can't expect nonprofits to do it themselves," urged one grant maker who believes in the value of outcomes evaluation. "You have to pay for it — but you shouldn't pay too much for it." The evaluations described by our interviewees ranged in cost from $35,000 to $300,000. We're back to proportionality again — proportionality with regard to the size of an organization's overall budget, the size of the grant, the maturity of the field, and the weight of the issue or problem.

Opinions differ on the need for consultant assistance. Many grant makers find it essential to have outside, impartial eyes and additional, trained staff assistance in the development and implementation of an evaluation. One experienced grant maker, though, said that "people make a big deal about independent review, but you can do it yourself."

Everyone agrees that multiple points of view are essential: client groups, advisory groups, colleagues. There is agreement, too, that the grantee must lead the process and use it to strengthen its internal capacity for self-evaluation.

Measurement matters: Seven tensions

If measuring program outcomes makes so much sense, then why is it controversial? As one experienced evaluator explained, "measures are not benign; they're tools" — and tools tend to get used for particular purposes. "Virtually every indicator we use in society is highly controversial," he continued, "from the consumer price index to the unemployment rate."

The debate over measurable outcomes is actually made up of several lower-level questions. Each explores a very real tension in the work.

1. Which needles are we trying to move?

It might appear that selecting which outcomes to measure is a straightforward proposition, but considerable nuance can attend the task. The selection process should pay careful attention to unexamined assumptions.

A director of a community organization described a project in which young people worked alongside social scientists to assess programs in juvenile justice. To measure the effectiveness of the programs, one of the outcomes the young people wanted to test for was "respect" — whether adults afforded it to program participants, whether young people had it for one another. Hard to operationalize, and perhaps more obviously interpreted as an input, such an outcome had not been considered by the sociologists. In the end, respect became an organizing principle for the entire evaluation. Through self-reporting surveys and interviews with both youth and adult staff from the criminal justice system, respect was revealed to be a critical variable: when it was missing from the program, there were major problems; when the program worked to foster respect, it was helpful. "The youth were trying to put a square peg in a round hole," said the director. "We had to make a square hole."

Another example was given by a grant maker whose foundation was building an evaluation into an existing portfolio of grantees, all of whom were engaging residents in solving community problems through civic participation. For the team that was developing the evaluation (it included grantees, program officers, and foundation evaluation personnel), the first job was to name the goal of the program, or the overall outcome being sought. The director of evaluation said the goal should be to improve selected conditions in the community, such as access to health care or student graduation rates. The program officer wanted the ultimate outcome to be "dense, interlocking networks of organizations that draw residents into efforts to improve community conditions." "I eventually realized," the program officer reported, "that, if the goal was to move the needle on specific problems, then the civic participation aspect was just a means to that end. Someone later could say, 'I've got a better approach for moving that needle.' But civic participation was not only a means but the very end that the grant making was trying to stimulate." As it turned out, the evaluation stressed the effect of civic engagement on participants' development, by looking at, for example, changes in civic skills and knowledge, level of participation in civic life, and self-identity as a leader.

2. Are we searching under the streetlamp?

The sociologists in the juvenile justice evaluation were hesitant to include respect as an outcome because they didn't know exactly how to measure it. "There is no consistent data set," explained a nonprofit executive, "on empowerment, or voice, or positive relationships. When the data don't exist, they can get washed out of the picture."

Scale — in both size and time — can be part of the issue. The larger the size, the more an evaluation will tend to rely on standardized sets of outcomes. There will be greater difficulty, said a community leader, in accounting for the diversity of local, indigenous knowledge, in getting to important detail in outcome measures. A major program to standardize arts education using outcomes approaches, for example, included almost nothing about creativity in its list of measurable skills. Difficult to assess in any case, creativity was much too hard to measure at the scale of the project.

Time scale, too, can drive measurement. The evaluation of the farming initiative mentioned above, for example, had initially set as a major outcome an increase in the number of acres of farmland managed with sustainable approaches. But for the time scale of the evaluation, data on acreage were hard to collect: they vary from year to year as crops are rotated, for one thing. So the evaluation settled on the next available proxies: the numbers of farmers getting certified in sustainable approaches and joining sustainable farming associations. Clearly some distance had been created between desired results and the measurement of them.

It is important to distinguish between the proxies of measurement — the "indicators" — and the actual changes you are trying to make. Incorrect selection of proxies can leave you with little helpful information. A major foundation initiative to reduce gang violence, for instance, settled on the proxy measurement of reduced youth homicides. Many millions of dollars and several years later, the evaluators realized that this measure would never respond to the initiative's interventions. Very little could be said about the effectiveness of the project.

Pursuing available proxies too slavishly can lead to real distortions. In the words of one grant maker, "if all you do is stuff you think you can measure, you are actually lowering the bar." "It's like the old joke about the drunk looking for his lost car keys under the streetlamp," said another. "When a passerby asks him why he's not looking on the next street, where he lost the keys, the drunk replies that the light is better here. You've got to be careful that you are not limiting yourself to the spot under the streetlamp."

3. Are we mistaking partial measures for the full truth?

A related danger is that indicators can become confused with actual outcomes. The ability to measure the proxy can entice you into thinking you've captured the truth. As the evaluation director of a major foundation explained, "There's a temptation to believe that you've got a full description of everything that's important. It can be exquisitely seductive, and it can be spurious."

"When outcome measures became the way for government to evaluate programs for runaway teens," recalled a nonprofit director, "there was a shift from process objectives, like beds being filled at night in shelters, to outcomes — family reunification, in particular." That sounded like a great idea, until it became clear that the metric was encouraging agencies to return some adolescents to abusive situations. The indicator was a measure of a partial truth.

Partial truths can be deadly. "Both space shuttle disasters were a result of a focus on outcomes that were too narrow," said a consultant evaluator. "Managers felt that getting the launch done was the most important outcome. The information about the faulty systems was known; the engineers were trying to get the information up the line to stop the launches."

The real point here is this: indicators are only that, *indicators* of effectiveness, and they can lead programs astray. "Suppose you've got an anti-smoking program," explained an evaluator, "and you're trying to figure out its effectiveness. You look at whether cigarette sales go down in the county. But are people buying outside the county? Your indicators might say nothing at all about this."

4. Can the push for accountability corrupt the data?

When evaluation becomes too heavily weighted toward accountability, argue leaders in the field, measurable outcomes can not only mislead, but corrupt. People will teach to the test. "What gets measured gets done," said one evaluator.

In a famous essay, "The Experimenting Society," Donald T. Campbell, one of the pioneers of modern evaluation of social programs, cites a litany of such abuse: the U.S. war on crime between 1968 and 1971 got results by downgrading the seriousness of recorded incidents; in the Soviet Union, nail factories overproduced large spikes when quotas were set by tonnage and small nails when quotas were set by number. More recently, the outcomes-based accountability approach of the No Child Left Behind law has led to cheating by some principals and superintendents. The Houston school district showed a reduction in dropouts by changing how the statistic was calculated.

Grant makers often worry that excessive emphasis on accountability can lead grantees to over-promise intended results, under-report missed goals, and even pursue incentives that sabotage the real work. "One hectic day at the airport," said a grant maker, "when everything was running late and people were scrambling to get on planes, I watched an airline employee close the door and send off a plane that was not full. She was obviously being measured on the time of departure. Maybe there was another perspective through which that could be seen as the best outcome, but I hoped I'd never done that to a grantee."

Said a former foundation executive, "We funders keep wanting to push the bar up, and grantees are understandably hesitant to object. It should be like the high jump: set the bar at a level people can get over. Then we can raise it over time."

5. Is it possible to measure impact meaningfully, given the realities of time and control?

"Time is a major issue," said the executive director of a program that works with high school students. "We measure impacts within a program period, but the real story won't unfold until the kid gets into college or not, stays in college or not," or even longer.

Almost every outcomes evaluation faces some version of the problem of aiming for long-term change but needing earlier indicators of success. "The tensions appear immediately," said one evaluator. "When do outcomes become visible? Later. But we need information earlier in order to strengthen the work. An awful lot of the intellectual effort goes into articulating short-term outcomes inspired by a long-term vision."

One famous story concerns evaluations of the federal Head Start pre-school program. An early study showed that the gains children made through the program disappeared after two to four years. But a more recent evaluation tells a different story. "Forty years later," explained the director of evaluation at a major foundation, "the Head Start participants are statistically different. What looks like it washed out at one period of time has turned out to be very important later down the line." Longitudinal studies are often prohibitively expensive, though — and of course they take time that is often simply not available.

Another thorny issue is the amount of control a particular intervention has over the results it seeks. This is partly a question of attribution — a question that can be answered definitively only with a big, very, very expensive control-group and randomized-assignment evaluation. But the question plays out in the relationship between grantees and funders.

The executive director of the youth program mentioned above explained that his organization is held accountable by the city council — the program's funder — for outcomes over which the school district actually has much more control. The city council knows this but has no formal control over the schools. "Sometimes," he noted ruefully, "those who end up accountable have the least power." A version of the problem can be found when philanthropy plays the role of the city council. "It is easy for us to expect too much of individual grantees," said one foundation official. "We are sometimes guilty of passing the buck by asking for results that are too big for any particular grantee."

6. Can individual measures take account of the full complexity of the environment?

Grant makers, grantees, and evaluators alike worry that a measurable outcomes approach can be overly simplistic. As a former foundation president said recently, "realities of social action and social change lend themselves only in a very clumsy way to the tidy world of numbers and bottom lines." That view certainly has opponents, particularly among grant makers who have worked to develop measures of "social return on investment" (see page 6). Yet the issue is there to be examined. "We presume that we know what outcomes to measure in social change work, but we don't," argued an evaluator. "We use models from medical and agricultural fields. But we don't have any idea why, for example, violence happens. These things are not linear and mechanical."

A consultant who coaches foundation staff on evaluation took a similar line of argument. "Social systems are a lot like coral reefs. They have a rich and complex structure of branching and budding, and to measure them we need to look at issues of growth, strength, subtlety, diversity."

Asked to reflect on the problem, one community leader explained, "Much of grant making is field building. The nonprofit world is an ecosystem with lots of necessary diversity, and there are many reasons to make grants other than the specific outcomes any particular grantee will produce. I was on the board of an organization that made $700 grants to public school teachers so they could innovate in their classrooms. Were we going to hold them accountable for outcomes? We were trying to help teachers have heart and motivation for the year; we wanted them to feel hope. Would we put in an outcome measure for 'teacher feels happier'?"

7. Do outcome measures help us learn about the things we want to understand?

At its heart, the debate often seems to be about whether or not it's possible to measure everything of value to philanthropy. Here is the way one grant maker describes it: "There's a difference between writing a business plan and raising a family. With a business, you've got that bottom line, which you want to maximize, and you plan backward from it. In raising your children, there are so many bottom lines! So many interventions! How many of them are quantifiable?"

An executive director of a community organization struck the same theme: "Nonprofits change people, and we are so ignorant about people. Suppose you have a program to increase gender balance in engineering. You give a fellowship to a young woman to study engineering, and halfway through she switches her major to art history. This is not an outcome failure of the college or the fellowship program. You could only change it by changing admissions to the program. But you have to take the risk."

Many interviewed for this guide said that the way to take the risk is to shift from an accountability focus — the "game of gotcha," in the words of one experienced grant maker, or, from an evaluator, "the current push to tote it up at the end of the day" — to a learning focus. Don't throw out the outcomes. But use several of them, from several angles; keep them in perspective, and keep the overall ecosystem in perspective. Think of outcomes as "meaningful signs of progress or trouble," suggested one evaluator. Said another, "Think of them as information targets, rather than performance targets or accountability targets. Treat outcomes as measures for dialogue and understanding."

"It is all about investing in a learning process," said another grant maker. "Outcome measures are just one tool for developing meaningful ways to tell your story."

Outcomes Dos and Don'ts

Do …	Don't …
Strive for a culture of intentional learning	Use only single indicators
Foster a dialogue assisted by the data	Lose sight of larger context and systems dynamics
Maintain flexibility for adaptation in a process of continuous feedback	Let the stakes get too high
Make it a learning experience, not an accountability experience	Confuse the indicator measures with the outcomes or the full truth
Seek balance and proportionality	Expect results from long-term strategies for large-scale problems

ACKNOWLEDGMENTS

Pam David
Tom David
Eileen Foley
Gail Greely
David Hunter
Barbara Kibbe
Colin Lacon
Jonathan London
Jan Masaoka
Leslie Medine
Julie Lieberman Neale
Tatum Nolan
Michael Quinn Patton
Elizabeth Reisner
David Scheie
Bruce Sievers
Peter Tavernise
Melinda Tuan
Marc Zegans

Alternatives in Action
Marguerite Casey Foundation
Cisco Systems Corporate Philanthropy
Edna McConnell Clark Foundation
CompassPoint Nonprofit Services
Cricket Island Foundation
Walter and Elise Haas Fund
Policy Studies Associates, Inc.
REDF
Skoll Foundation
Stanford University
The Stuart Foundation
Touchstone Center for Collaborative Inquiry
Youth in Focus

Underwriting for this guide was provided by the Ford Foundation.

For additional guides and other materials in the GrantCraft series, see www.grantcraft.org

grantcraft
PRACTICAL WISDOM FOR GRANTMAKERS

www.grantcraft.org
A PROJECT OF THE FORD FOUNDATION