

What Impact?

A FRAMEWORK FOR MEASURING THE SCALE AND SCOPE OF SOCIAL PERFORMANCE

**Alnoor Ebrahim
V. Kasturi Rangan**

Organizations with social missions, such as nonprofits and social enterprises, are under growing pressure to demonstrate their impacts on pressing societal problems such as global poverty. This article draws on several cases to build a performance assessment framework premised on an organization's operational mission, scale, and scope. Not all organizations should measure their long-term impact, defined as lasting changes in the lives of people and their societies. Rather, some organizations would be better off measuring shorter-term outputs or individual outcomes. Funders such as foundations and impact investors are better positioned to measure systemic impacts. (Keywords: Performance Measurement, Scale, Scope, Mission, Nonprofit, Social Enterprise)

The world of nonprofit organizations, philanthropy, and social enterprise has been preoccupied with two powerful mantras in recent years. Since the early 1990s, the refrain of “accountability” has been ascendant, with funders, taxpayers, concerned citizens, and clients demanding that nonprofits be more transparent about their fundraising and spending, how they are governed, and what they have achieved with the resources entrusted to them.¹ A more recent manifestation of this discourse has centered on the mantra of “impact,” or demonstrating results in addressing complex social problems such as poverty and inequality.² This attention to impact, following on the heels of accountability, is mainly driven by funders who want to know whether their funds are making a difference or might be better spent elsewhere. It is also driven by an increasing professionalization of the sector, which has led to the emergence of common administrative norms including the use of credentialed experts such as auditors and evaluators.³

In this article, we identify key developments around impact measurement and their implications for performance management in social sector organizations. By social sector, we refer to organizations driven primarily by a social purpose,

The authors are grateful to the editor and three reviewers for their insightful suggestions, and for the support of the Division of Research and Faculty Development at Harvard Business School.

Alnoor Ebrahim is an Associate Professor in the Social Enterprise Initiative at Harvard Business School. <aebrahim@hbs.edu>

V. Kasturi Rangan is Co-chair of the Social Enterprise Initiative at Harvard Business School, where he is the Malcolm P. McNair Professor of Marketing. <vrangan@hbs.edu>

such as nonprofit or nongovernmental organizations and social enterprises. We offer a framework for determining which kinds of measures are appropriate, as driven by the mission and goals of the organization. Our normative argument is that it is not feasible, or even desirable, for all organizations to develop metrics at all levels of a results chain, from immediate outputs to long-

term societal impacts. The more important challenge is one of alignment: designing metrics and measurement systems to support the achievement of well-defined mission objectives. Measurement efforts extending beyond this frame are a distraction from the operational work of the organization. Often, it is the funder, who sits at a higher level in the social sector ecosystem, that will have a broader and more integrative perspective on how the work of several implementing organizations fits together to advance systemic goals. Such impact is better measured at the funder level, leaving the individual organizations to do what they do best—focus on their more specific missions.

The Current State of Social Performance Measurement

Much of the literature on the topic of performance in the social sector is under-theorized and in need of conceptual framing. As a result, the bulk of the material we cite below is not from academic journals; instead, it is taken directly from practice—from influential think tanks, government agencies, funding organizations, and prominent consulting firms that are shaping the practice of social performance. The academic literature in nonprofit studies, philanthropy, and management lags behind in providing theoretical and analytical insights to this burgeoning field (see Appendix 1 for a summary of the diverse approaches to social performance measurement). Second, because of the vastness and diversity of the social sector, we focus our discussion on organizations working broadly in the field of human services, public benefit, and poverty alleviation. In 2010, there were roughly 300,000 such organizations in the United States alone, accounting for roughly \$500 billion in revenues. We leave out a discussion of organizations primarily engaged in advocacy (such as human rights and environmental policy organizations) or the arts and culture (such as museums, symphonies, and dance companies), where we believe the measurement of performance is even more complex and nuanced.

The most widely advocated set of approaches to social performance measurement involve an assessment of impacts or results, which are broadly labeled as “impact evaluation” and “outcome measurement.” Such assessments are often conducted after program implementation is complete and are driven primarily by funders such as foundations and governments: Has the program reduced poverty? Has it improved health outcomes? Has it placed trainees in permanent jobs, and thereby increased their lifetime earnings potential? They involve a range of methodologies drawing on benefit-cost analyses, summative evaluation, and experimental methods such as randomized control trials.

The term “impact” has become part of the everyday lexicon of social sector funders in recent years, with frequent references to “high-impact nonprofits” or “impact philanthropy” and “impact on steroids.”⁴ But the term impact has not been consistently defined. An established literature in international development and evaluation often uses the term to refer to “significant or lasting changes in people’s lives, brought about by a given action or series of actions.”⁵ More recently, impact has also come to be associated with results that target the “root causes” of a social problem.⁶ Others use impact more narrowly to refer to an organization’s specific and measurable role in affecting a social result (attribution) requiring a counterfactual for assessment.⁷ In this article, we distinguish between outcomes and impacts, with the former referring to lasting changes in the lives of individuals and the latter to lasting results achieved at a community or societal level. Attribution remains a challenge regardless of whether one is assessing outcomes or impacts.

Many frameworks for measuring social performance employ a “results chain” or “logic model,” which has its roots in the evaluation of programs and projects, and was originally developed for the United States Agency for International Development (USAID) in the late 1960s.⁸ Table 1 shows the key components of the basic logic model—inputs, activities, outputs, outcomes, and impacts—and includes examples of the types of measures typically included under each step. Through their adoption and dissemination by program evaluation specialists, logic models have emerged as a primary means through which social sector organizations identify impacts and other performance metrics.

Funding organizations—from philanthropic foundations and governmental agencies to impact investors—increasingly expect the organizations they support to measure their outcomes and impacts. However, the evidence on whether outcome measurement has led to improved performance is mixed. A study of thirty leading U.S. nonprofits found that measurement was useful to the organizations for improving outcomes, particularly when they: set measurable goals linked to mission (rather than trying to measure mission directly); kept measures simple and easy to communicate; and selected measures that created a culture of accountability and common purpose in the organization, thus helping to align the work of disparate units and chapters.⁹ The United Way of America, which was one of the first national agencies to ask members of its network to distinguish between outputs and outcomes, showed gains in effectiveness among 391 agencies that it surveyed in 2000. It reported that an overwhelming proportion of its partner organizations found outcome measurement useful for communicating results and identifying effective practices (84-88%), as well as for helping to improve service delivery of programs (76%). However, the results were not all positive: a significant number of agencies reported that implementing outcome measurement has led to a focus on measurable outcomes at the expense of other important results (46%), has overloaded the organization’s record-keeping capacity (55%), and that there has remained uncertainty regarding how to make program changes based on identified strengths and weaknesses (42%).¹⁰

Foundations, too, have a mixed record of using impact assessments and evaluation in their decision making. In an analysis of evaluation methods used

TABLE I. Logic Model

Inputs	→	Activities	→	Outputs	→	Outcomes	→	Impacts
<ul style="list-style-type: none"> ▪ funds ▪ equipment and supplies ▪ knowledge and technical expertise 		<ul style="list-style-type: none"> ▪ basic needs delivery, such as food and shelter ▪ service delivery, such as job training and counseling ▪ infrastructure construction, such as transportation 		<p>Results: immediate</p> <ul style="list-style-type: none"> ▪ people fed, housed, or treated ▪ people trained or educated ▪ roads built and goods transported to market 		<p>Results: medium- and long-term</p> <ul style="list-style-type: none"> ▪ improved quality of life, health, educational attainment, etc. ▪ increased incomes <p>(measured for individuals)</p>		<p>Results: effects on root causes, sustained significant change</p> <ul style="list-style-type: none"> ▪ sustained drop in poverty (or obesity, illiteracy, etc.) ▪ improvements in human development indicators <p>(measured in terms of communities, populations, or ecosystems)</p>

by philanthropic foundations over three decades since the early 1970s, nonprofit historian Peter Dobkin Hall has argued that such evaluations lack rigor, and that key decision makers are often indifferent to the findings.¹¹ Others have suggested that philanthropic giving is often motivated by the expressive interests of donors and not necessarily by evidence of what works and what doesn't.¹² Many foundations continue to struggle with how to integrate a range of measurement approaches into their decision making. A 2007 survey of emerging approaches to evaluation in philanthropy pointed to a "shift from the use of evaluation to measure the impact of past grants and toward a more timely and pragmatic process of gathering forward-looking information that will enable both grantors and grantees to make ongoing improvements in their work."¹³ The report found this tension to be fairly common in foundations, motivated on one hand by a need to identify the long-term outcomes and impacts of past projects, while also seeking more flexible and timely sources of data to help with present performance.

In recent years, however, there has been considerable progress in developing measurement and evaluation methods with numerous approaches being developed by prominent consulting firms (see Appendix 1), as well as a more deliberate emphasis on measurement by foundations and impact investors. These include not only established philanthropic institutions, such as the William and Flora Hewlett Foundation, the Edna McConnell Clark Foundation, and the Rockefeller Foundation, but also a number of relatively young entrants to the funding field—The Bill and Melinda Gates Foundation, Robin Hood Foundation, and Acumen Fund, to name just a few—that have advocated for an explicit emphasis on measurement and the development of standardized metrics.¹⁴

At the same time, a chorus of skeptical voices, particularly from practitioners, has suggested that while impact and outcome measurement appears to be "a good tool to help funders see what bang they're getting for their buck," it runs the risk of being counterproductive in the long run, both by drawing precious resources away from services and by putting too much emphasis on outcomes for which the causal links are unclear, thus reflecting more of an obsession with institutional expectations of accountability to funders than an interest in actually finding ways of improving services and results.¹⁵ Other practice-focused studies have also concluded that in the end, a project's success had less to do with whether measurement systems were developed and more to do with "whether the organization was able to create a culture that valued the process of self-evaluation."¹⁶

In short, this practitioner-based literature suggests that outcome and impact measurement are undertaken for multiple purposes, including evaluating effectiveness, satisfying external accountability expectations, and guiding organizations in improving their actions. However, for all the intended benefits, their adoption does not necessarily improve outcomes, impacts, or decision making.

What to Measure? A Performance Framework for Operating Organizations

Conventional wisdom in the social sector suggests that one should measure results as far down the logic chain as possible, to outcomes and societal impacts.

This expectation is based on a normative view that organizations working on social problems, especially if they seek public support, should be able to demonstrate results in solving societal problems. Yet it is worth considering whether, and to what degree, such measurement makes sense for all social sector organizations.

For instance, many international NGOs—such as the Red Cross and Doctors Without Borders—are engaged in emergency relief work. Measuring the work of such organizations is conceptually fairly straightforward: count the timeliness and delivery of emergency supplies such as tents, food, water, and medical supplies, as well as the numbers of people reached. Emergency relief is thus typically measured in terms of activities and outputs. While it is a complex activity, requiring highly sophisticated coordination and logistics management capabilities, it is focused on meeting immediate survival needs rather than long-term development outcomes or impacts. The links between inputs, activities, and outputs follow logically: the organization plans its requirements of supplies and staff (inputs) and the logistics for delivering those supplies (activities) in order to provide relief to the people most affected by the emergency (outputs). When the effort is well planned and executed, the program will be able to orchestrate activities that lead to measurable outputs.

Outcome measurement, on the other hand, requires answers to a more complex causal question: Are the activities and outputs leading to sustained improvements in the lives of affected people? Outcome measurement is less common and more difficult to do, given that organizations have the most control over their immediate activities and outputs, whereas outcomes are often moderated by events beyond their organizational boundaries. For example the emergency relief organization that has done excellent work during and after a natural disaster might still fall short on outcomes of rehabilitating and resettling those displaced from their homes and livelihoods, especially if those outcomes depend on extended coordination with local governments, businesses, and other NGOs. Connecting outcomes to societal impacts, such as a sustained drop in poverty in the region, is even more complex due to the number of additional factors at play—involving the larger political, social, cultural, and economic systems—that are beyond the control of any one entity. In short, outputs don't necessarily translate to outcomes, and outcomes don't necessarily translate to impact.

This much is clear: every organization should at least measure and report on its activities and outputs, as these results are largely within its control. However, when should it step forward into the domain of outcomes and impacts? Two organizations that have confronted this problem in different parts of the world are Aravind Eye Hospital in southern India and the Harlem Children's Zone in New York City.¹⁷

In 2012 alone, Aravind Eye Hospital performed over 340,000 surgeries, most of them for cataract surgeries in one province in India; and it screened over ten times that number of people. Its outputs are remarkable: providing vision correction to over 3 million individuals since its founding in 1979. However, Aravind realizes that output alone, judged by the number of patients screened or treated, is not an adequate indicator of its performance. Rather it needs to know the outcomes of its intervention. How many people who received treatment were

satisfactorily cured? In other words without a high quality intervention, its outputs do not convert to outcomes. Aravind engineered a superb operational process by which the surgery was performed at the highest quality standards. The rate of complications year after year were less than half that of hospitals in the United Kingdom.¹⁸ With such a high quality of service, Aravind was able to establish a tight linkage between outputs and outcomes. We may never know whether those outcomes resulted in poverty reduction (impact), but it is logically reasonable to assume that those with recovered eyesight would be better able to fend for themselves. Even given the tight linkage in Aravind's operations between outputs and outcomes, the organization assumes but does not measure impact—that individuals with recovered eyesight from cataract treatment will be able to lead productive lives once again and thereby contribute to society. While this assumption seems reasonable, the organization has cautiously stayed away from making that leap and seeking to take credit for impacts.

In contrast, there is Harlem's Children Zone (HCZ) in New York. While Aravind is highly focused in its scope of activities, conducting vision screening and cataract surgeries, HCZ offers a broad scope of activities aimed at developing a child's capabilities through various stages of educational development. The organization runs pre-natal programs for parents; early childhood and after-school programs for children; charter schools, vocational training, and college preparation for youth; health programs to help families address asthma and obesity; and even programs for adults that aim to create a supportive and caring community for children. Unlike Aravind which focuses on a singular intervention, HCZ attempts to provide a "pipeline" of interventions "from cradle to college to community" that collectively enhance the chances of a child in Harlem making it from school to the workplace and on to self-sufficiency and responsible citizenry. HCZ's "theory of change" is that educational support has to be continuous from pre-school through high school, and it has to be supplemented by extracurricular and community support to solidify the young person's all-round development.¹⁹

In order to provide this pipeline of comprehensive interventions, HCZ has concentrated its activities in a narrow geographical region of nearly 100 city blocks of Harlem, under the assumption that it will be better able to control the child's overall environment. In 1998 before HCZ scaled its programs across Harlem, 60% of the children lived in poverty and only 20% of children in elementary schools were able to read at grade level. In 2011, the sixth graders in its two main charter schools had shown significant improvements: approximately 80% were at or above grade level in statewide math exams, and between 48-67% (depending on the school) were at or above grade level in English. Moreover, 95% of seniors in public schools who attended HCZ after-school programs were accepted into college.²⁰ The grade level metrics are primarily output measures, while college acceptance may be considered an outcome measure. The time horizon for these interventions is long (5 to 19 years), and the organization is undertaking longitudinal studies to better assess its results. Even then, drawing a causal link between HCZ's interventions and longer-term outcomes such as lifetime incomes of its graduates, and impacts such as a decline in poverty in Harlem, remains complicated due to numerous social and economic factors that HCZ cannot control.

A key distinction between Aravind and HCZ is how they make the leap from outputs to outcomes. Aravind can measure outcomes because of the tight causal linkage between its outputs (corrective surgery) and outcomes (quality vision), which hinges on the quality of its surgery. HCZ can measure outcomes because it is able to vertically integrate a comprehensive set of interventions (a pipeline from cradle to college) that it controls in a tightly bounded geographic space (100 city blocks). More generally, measuring outcomes is possible under two conditions that are uncommon in the social sector: when the causal link between outputs and outcomes is well established, or when the range of the integrated interventions needed to achieve outcomes are within the control of the organization.

Clarifying the Operating Mission

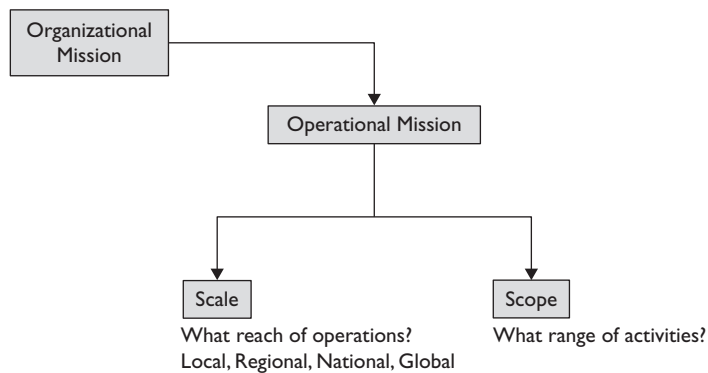
These preliminary observations lead us to the construction of a framework that relates performance measurement to the mission of a social sector organization. A unique characteristic of such organizations is their drive towards mission fulfillment, which is their primary purpose. Implicit within an organization's aspirational vision, which may be eliminating blindness or hunger or poverty, is a much more modulated "operating mission" that focuses on the immediate work at hand.²¹ From an operational perspective, Aravind's mission centers on providing surgical correction to those suffering from cataract blindness, regardless of their ability to pay. Knowing that nearly 75% of the blindness in India is caused by cataract formation, starting with such a sharp focus helps the organization make rapid progress against its mission. HCZ's operational mission is to ensure educational outcomes for kids in Harlem. Even though the organization aims at a larger level to address poverty in Harlem, its day-to-day activities and programs concentrate on improving the quality and context of children's education.

Thus while a nonprofit or social enterprise may have an aspirational mission for what the world should look like, in practice its work is best captured by its more pragmatic operational mission. This operational mission is not always explicitly articulated, but it can be observed in the work that the organization attempts to do on a day-to-day basis. Figure 1 depicts this distinction. The operational mission, and how to measure progress towards achieving it, can be further understood by examining the scale and scope of the organization's work.

Scale and Scope

Implicit in every social mission is a statement of the scale of the problem the organization intends to address and the nature of the intervention required to address it. While Aravind's general mission makes reference to the population of blind people in the world, its operational mission is much more narrowly focused on one provincial region, Tamil Nadu State with about 750,000 blind people, and the surrounding region with roughly another 2 million blind people. Nearly half of this group is poor and can't afford to pay for treatment. In operational terms, this is Aravind's target client group against which performance should be judged. For HCZ the target area is narrower, the 100 city blocks covering Harlem, and the nearly 20,000 children who attend Harlem's public schools.

FIGURE 1. Operational Mission



The scale of an organization’s operations can be expected to evolve with time. As it gains reputation and funding, the organization will be attracted to expand the reach of its operations. Consequently performance assessments must be based on the explicit target identified by the organization in its operational mission. Aravind started off in one city, Madurai in south India, and then steadily expanded through outreach programs and more hospitals to blanket its home state, eventually branching out to neighboring states. By 2012, it was performing nearly 350,000 surgeries in its focused region. Harlem Children’s Zone started with 24 blocks in central Harlem in 1997, steadily expanding to cover the entire 100-block region within a decade, reaching over 9,400 children and 6,200 adults within this urban zone in 2011. Thus Aravind seeks to address the problem of vision at a regional scale, while HCZ tackles the problem of educational attainment at a local scale. For these two organizations, it makes sense to measure performance against these regional and local operational targets, respectively.

The second dimension, scope, is a measure of the range of activities required to address the need identified in the operational mission. As noted above, the scope of Aravind’s intervention is narrow, focused on providing high quality eye surgery. The organization can measure outputs (number of people screened and treated), which are then easily and explicitly linked to outcomes (number of people with quality vision). Figure 2 illustrates this explicit link from outputs to outcomes in the scope of Aravind’s activities (the darker arrows). In the last decade, Aravind has taken advantage of its vast outreach (screening 3 million individuals) to add a second line of treatment—fitting eyeglasses for vision correction. This addition in scope is aligned with its mission: like cataract surgery, once individuals receive corrective lenses, their vision improves. Again, the causal link between outputs (number of people receiving glasses) and outcomes (quality vision) is explicit, provided that quality is maintained in the fitting of eyeglasses. The case of Aravind provides an example of a narrow scope of activities resulting in measurable outcomes. Do corrective surgeries or lenses lead to an improved quality of life and reduced poverty for affected people (impact)? Although this

link between outcomes and impact seems highly plausible, Aravind does not attempt to measure impacts, preferring to stick to what it does best: delivering quality surgery and eyeglasses (outputs), which are tightly linked to the outcome of improved vision. The link between outcomes and impacts is implicit in its work (hence the lighter arrow in Figure 2). Aravind’s performance can thus be assessed in terms of its effectiveness in producing quality vision at a regional scale.

HCZ’s performance can be similarly examined in terms of scale and scope (see Figure 3). Its scale is local, centered on the zone in Harlem where it operates. However, its scope is very wide, made up of activities from early childhood education to college preparation. As children pass through its pipeline, HCZ hypothesizes that the combination of outputs (such as educational experiences, test scores in math and English, conflict resolution, and parenting skills) will add up to measurable outcomes in the form of college acceptance and quality jobs (the darker arrows in Figure 3). Some of these links are substantiated through research, while others are hypothetical. HCZ is able to track outcomes because its services to children are vertically integrated all the way to job placement and college admission. But even if the outcomes are achieved (college and jobs), will they lead to impacts such as reduced poverty in Harlem? This link is implicit in the organization’s work (the lighter arrow in Figure 3), although the longitudinal studies that it is undertaking may help to tighten the causal connections and make them more explicit. In short, HCZ’s performance can be assessed in terms of its effectiveness in producing educational outcomes for children at a local scale. The time horizon for these outcomes is 19 years.

In sum, the notion of scope captures the set of activities necessary for addressing a social problem, while scale captures the target size of the problem. The problem itself is articulated in the organization’s operational mission. Clarity on all three components—operational mission, scale, and scope—is necessary in order to know what to measure. Specifying the expected time horizon for results

FIGURE 2. Aravind’s Performance Vector

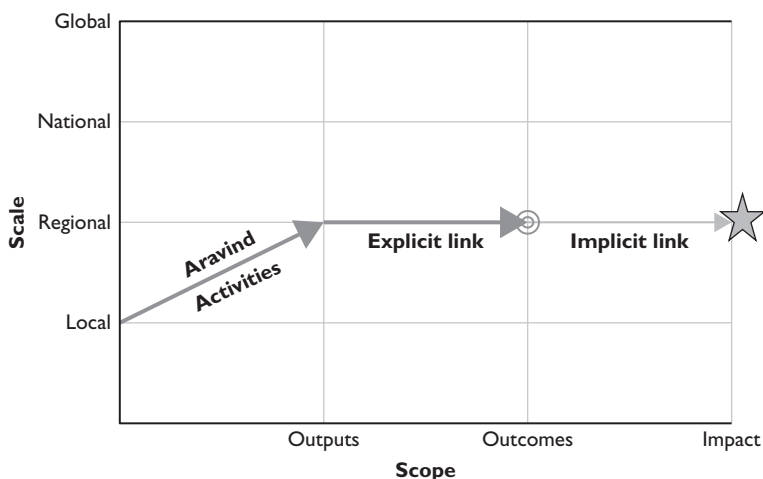
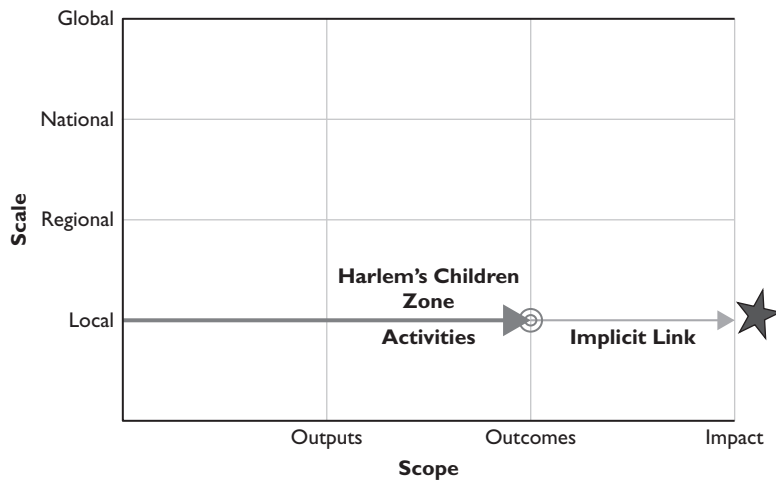


FIGURE 3. Harlem's Children Zone's Performance Vector



is also important, particularly where outcomes and impacts are expected to be visible only after many years.

Improving Performance: Expanding Scale and Scope

Our performance framework suggests that social sector organizations should primarily focus on delivering against their operational mission. Clearly, all organizations should be capable of measuring the outputs of their operations. However, only some will be able to go further to make credible and measurable claims about outcomes. This is possible under two conditions:

- the organization implements a narrow scope of activities where the causal link between outputs and outcomes is clearly established through evidence (e.g., Aravind’s eye surgeries lead to improved vision); or
- the organization implements a broad scope of activities that is vertically integrated to increase control over outcomes (e.g., HCZ’s cradle to college pipeline).

The time horizons for achieving outcomes can vary considerably. While Aravind may be able to gather evidence of improve vision fairly soon after surgery, HCZ’s interventions require a decade or more to see their outcomes materialize. Only rarely will organizations be in a position to go even further by claiming long-term sustained “impacts” on their communities and society. These two conditions further suggest how organizational performance can be improved: by expanding operations, while retaining the current scope, in order to reach a larger target population (increase scale); and/or by offering more services down the logic chain that will enable greater control over outcomes (increase scope). Whether such an expansion of scale or scope is warranted is a strategic question for the organization’s leadership and key supporters.

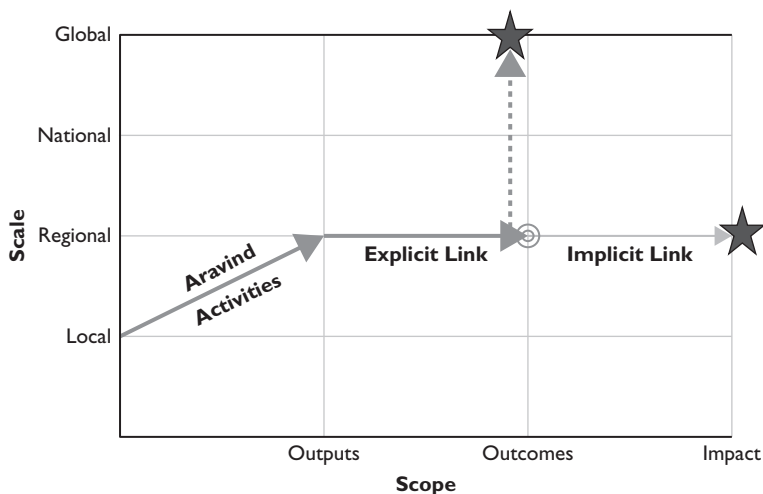
Scale in the social sector can be achieved not only through organizational growth, but also via a myriad of other means, particularly through influencing public policy and coalition building, training others to replicate and adapt its model, or even through the creation of new industries.²² For instance, nonprofits served as the initial pioneers in the microfinance industry, but the explosion of the industry was fueled by the creation of a supportive regulatory environment and the entry of opportunistic private sector players.²³

Expanding scope in the social sector, in order to increase control over outcomes, is not limited to vertical integration within the organization. An alternate strategy is to partner with organizations that carry out complementary work along the results chain. For instance, Aravind often gets asked to participate in blindness prevention programs. At a superficial level, such an upstream expansion in scope might seem to meet the stated mission of the organization, but when examined more closely does not fit the capabilities of the organization. Prevention programs require educating and persuading the community, especially households, to alter the way they prepare and cook their food, such as with locally available plants and vegetables with high content of Vitamin A. The target population is families with children, not those who already are blind. Moreover Aravind's core operating model is centered on delivering high quality surgical intervention rather than on mass-media prevention campaigns. To expand scope, Aravind would be better off partnering with a social marketing organization that could fill a distinct niche rather than integrating this new activity into its existing organization.

Aravind has chosen to dramatically expand the scale of its operations but not its scope. First, it has launched a training facility where hundreds of administrators and healthcare workers from 43 countries come to learn the Aravind "management model." They learn about its low-cost and high-quality patient care systems, its outreach programs, and its hybrid funding model (roughly 35% of the patients pay market prices and the rest of the 65% pay below market prices, with many getting free treatment). Armed with this knowledge they go back to their regions or countries to launch their own version of the Aravind model. In certain chosen locations, Aravind sends a team of doctors and nurses to actively engage with the local sponsor to set up systems and ramp up the model. For example Aravind's start-up team helped Grameen Bank to launch an eye hospital in Bangladesh. Such an approach scales the work of Aravind through virtual networks, as depicted by the dashed arrow in Figure 4.

On the other hand, if Aravind sought to expand its scope in order to have greater impact, it might partner "upstream" with organizations working on prevention and "downstream" with organizations that help its clients gain entry into the workforce once their vision is restored. Notably, both options for improving performance—expanding scale or scope—can be achieved by collaborating with other organizations rather than by attempting to grow the organization. Even HCZ, which has steadily expanded the scope of its activities through vertical integration across the results chain, remains local in scale. Any expansion in scale to a national or regional level will likely require collaboration with other organizations that could replicate its model in other cities. The scaling challenge for HCZ lies in

FIGURE 4. Expanding Scale and Scope



identifying the most critical components of its results chain and how those might be reproduced elsewhere. More generally, when both scale and scope have to be addressed, we expect that it will be especially difficult for organizations to achieve performance single-handedly.

Performance and the Role of Funders

Performance measurement does not operate in a vacuum and is the subject of much tension between operating organizations and their funders. Most organizations are typically resource-dependent on their funders, be they philanthropic foundations, governments, donors, or investors. Much of the growing emphasis on measuring performance and impact in the sector has come from institutional funders seeking to allocate their resources more effectively and strategically²⁴ and in publicly legitimizing their allocation decisions.²⁵ Performance measurement thus plays a critical and highly politicized role in mediating the relationship between operating organizations and their funders, and in building capital markets that reward performance.

How do funders assess their grantees or investees? As noted, funders have a mixed record of using performance assessments and evaluation for making decisions about grantees. How do funders assess their own performance? While many often demand performance measurement from their grantees, rarely are they explicit about how they measure their own performance. The most critical challenges of performance measurement lie not at the level of operating organizations, but among aggregators such as foundations, governments, and impact investors. It is at this level—where the funder is able to oversee hundreds of operating organizations—that it is possible to measure societal impacts.

What Funders Measure

We examined three funding organizations widely considered by their peers to be leaders and innovators of performance measurement: Acumen Fund, Robin Hood Foundation, and Millennium Challenge Corporation (MCC).²⁶ All three provide capital to operating nonprofits, social enterprises, or government agencies, either as investors or grant makers.

What do these organizations seek to measure? It turns out that they assess different things: Acumen measures outputs, such as mosquito nets made and distributed; Robin Hood focuses on long-term outcomes in the lives of individuals, such as gains in income; and MCC aims for both individual outcomes and broader impacts on society, such as reduction in poverty rates.

Consider Acumen Fund, a **venture philanthropy firm** with a portfolio of over 75 investments in for-profit social enterprises in Africa and Asia. Its primary social metric is the number of lives reached in base-of-pyramid markets (an output measure). **This means that when it invests in a company that manufactures anti-malarial bed nets, Acumen will count the number of nets manufactured and distributed.** For an enterprise that builds toilet and shower facilities in urban slums and business districts, it will tally up the number of times the toilets and showers are used. **However, why stop at these output measures, rather than outcomes such as reduction in malaria or improvements in health and environment?** Acumen's former chief information officer argues that doing so is complicated, expensive, and often impractical for early-stage enterprises.²⁷ **Instead, Acumen's strategy is to review the literature and consult experts to establish a link between a specific output and impact (for instance, how bed net distribution leads to malaria reduction) and to then count the outputs. This approach is similar to Aravind's, where the emphasis is on measuring outputs and their related outcomes, while using existing research or evidence to make credible claims about the link to community impact.**

Robin Hood, on the other hand, is a grant-making foundation created by hedge fund managers with a penchant for hard numbers. With a mission to fight poverty in New York City, the foundation puts each of its 200-plus grants to nonprofits through a cost-benefit analysis every year. One of its primary metrics is the expected increase in lifetime earnings of its clients. For grants focused on education, for example, its staff first identifies a set of results that can be immediately observed—such as school attendance, standardized test scores, and high school graduation. Then, they search for studies that link those measures to expected lifetime earnings or quality of life. For instance, some longitudinal research suggests that a 10% increase in test scores is correlated with a 4% increase in high school graduation rates, which in turn is associated with \$6,500 in increased income per year. Robin Hood uses these figures with caution, employing them as placeholders for estimating benefits until better research comes along. Initially the foundation placed the burden of measurement on its grantees, but quickly found that few grantees had the capacity or resources to do so. As a result, the organization has since built its own internal capacity for developing outcome metrics, employing a full-time researcher and contracting with a research firm for further support.

The MCC's approach is even more complex as it operates on a 20-year time horizon. The U.S. government agency makes grants to emerging market countries to reduce poverty through economic growth. Its extensive due diligence process—often two years long—first analyzes the barriers to economic growth in the country, and then identifies the sectors where the grants would most likely reduce poverty. For example, the MCC awarded \$547 million to the government of Ghana to build roads and ferries to get farm goods to market. To start, it estimated the number of farmers likely to benefit, and what those benefits would be: reduced cost and time of getting goods to market, access to new markets, and opportunities for wage employment. These data were used to anticipate an economic rate of return, with the primary outcome metric being increases in farmer incomes, along with impact metrics such as a reduction in regional poverty rates. Once a contract is signed, a monitoring and evaluation process kicks in to allow for mid-course correction. Benefits are expected to begin accruing only after five years, once the infrastructure is built and operating. The MCC plans to eventually conduct long-term evaluations, but has not yet done so. Unlike Acumen and Robin Hood who have influence on the process as it pertains to outputs and in some instances to outcomes, MCC's big bet on impacts on poverty depends on how the Ghanaian government and other complementary players build on its efforts to promote economic growth.

Among these three funding organizations, two features of their measurement approaches stand out. First, all three funders develop metrics in partnership with the operating organizations, trying not to overburden them. Acumen focuses on metrics that are useful to the entrepreneur in building the business. Robin Hood works with grantees to identify key outputs, but then elaborates the links to outcomes itself. And the MCC works with government agencies to build their capacities for implementing projects and tracking progress, often sending its own staff and evaluators to provide support. These kinds of partnerships are rare; more commonly funders expect operating organizations to measure outcomes and impacts themselves, without providing the funding or technical support to do so. Many expect accountability, but few are willing to support the overhead costs necessary for enabling it.

Second, none of these three funding organizations typically measures impact. They *hypothesize* what the outcomes and impacts might be but only in some instances are they able to follow through by commissioning their own research or multi-year evaluations. These are sophisticated funders and investors who are much better positioned to measure long-term results than the front-line organizations that contend with funding shortages and operational challenges every day. Surely measuring impact matters, but the experiences of these funders suggest that there are severe constraints to doing so. **Assessing impact requires a level of research expertise, commitment to longitudinal study, and allocation of resources that are typically beyond the capabilities of operating organizations and sometimes even their funders.**

Measuring the Performance of Funders

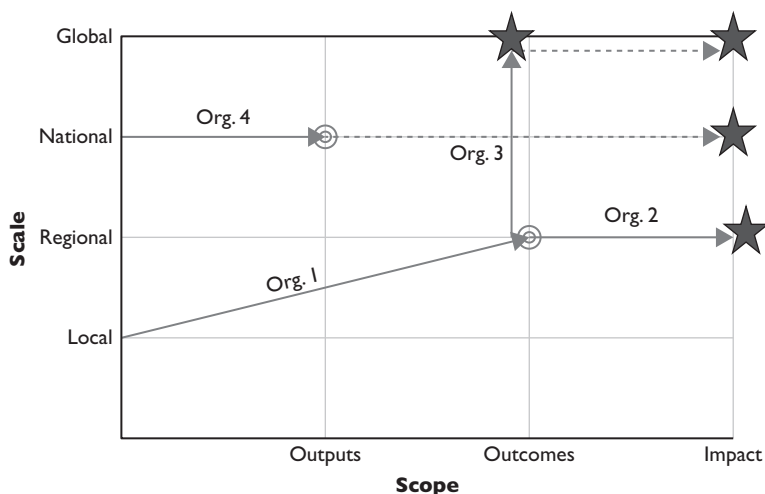
While funders such as foundations, impact investors, and governmental aid agencies seek to assess the performance of their grantees or investees, it is less

common for them to apply the same standards to measuring their own performance. Foundations that have large endowments face less pressure to do so, as they are not resource-dependent on their external environments. However, why would it matter for funders to assess their own performance, given that the actual implementation is done by operating organizations? Because it is at the level of the funder that systemic impacts—long-term sustained changes in society—can best be observed. Impacts are rarely achieved by individual organizations acting alone, but more often by collections of actors working towards a shared goal.²⁸ Because they support and oversee hundreds of nonprofits and social enterprises that typically act independently of one another, funders are specially positioned to connect that work in order to assess impacts.

A foundation that supports efforts in health care, for example, is uniquely situated to see how the work of all its grantees might link together as a portfolio—to connect the dots, or vertically integrate, among a series of outputs and outcomes in order to assess how they might lead to impacts. A central strategic challenge for funders is to articulate their own causal pathway for achieving impacts and to partner with an integrated portfolio of social sector organizations to achieve those impacts. Figure 5 illustrates this argument. The solid lines represent the activities and results of operating organizations such as nonprofits and social enterprises, which can be linked together by the funder (integrated) in order to achieve impact. The dashed lines represent the links between outputs, outcomes, and impacts established by the funder through its support of research that helps establish causal links.

It is unlikely that there is a single best way for funders to assess their own performance, or the collective performance of their grantees or investments. The Robin Hood Foundation calculates the average benefit-cost ratio for all of its investments. It estimates that for every dollar it spends on grants, the foundation

FIGURE 5. Systemic Impact: A Role for Funders



improves living standards for poor people by 18 dollars. This model is popular in the impact investment community, which seeks to measure the social return on its investments. Bilateral government agencies that support thousands of anti-poverty projects around the world are also investigating such aggregation methods, especially as they face pressure from taxpayers to demonstrate the results of international aid. A critique of this model is that it may atomize the social sector, creating a market that rewards the performance of individual organizations while discouraging integrative or collective action.

Acumen Fund works on a different model in which it attempts to identify promising organizations working on a social problem and helps them to scale. Its success can thus be measured in terms of the growth and reach of the organizations it funds. Acumen Fund invests in social enterprises that it believes can be scaled to reach over 1 million clients in base-of-pyramid markets. The organization provides not only “patient capital” over a period of several years in the form of debt or equity, but also offers management capacity building in order to help its investees build businesses capable of scaling.

While Acumen Fund attempts to create impact through scale, the Millennium Challenge Corporation attempts to do so along the scope dimension by assembling a complementary collection of several public agencies, private sector firms, and nongovernmental organizations in the service of its mission to alleviate poverty through economic growth. The Bill and Melinda Gates Foundation’s approach to global health can also be characterized in these terms, as the organization brings together key players needed to eliminate major infectious diseases. For instance, a \$100 million joint effort by Merck Pharmaceuticals and The Gates Foundation in Botswana aims to set up a comprehensive healthcare screening, testing, consulting, and delivery system for HIV/AIDS through a number of local and international partner organizations in the country.²⁹

Conclusions and Reasons for Optimism

The core of our framework for measuring social performance is relatively simple: clarify the operational mission, specify the set of activities to address that mission (scope), and identify the target size of the problem (scale). Yet, such measurement is rare in practice. Our approach provides a basis for assessing performance that is empirically rooted, rather than relying on unsubstantiated claims and unrealistic expectations about performance. It is a general approach that can be applied to any organization, while still allowing organizations to adapt their metrics of scale and scope to their specific contexts. All too often, social sector organizations seek to measure, or take credit for, impacts that extend well beyond the scale and scope of what they actually do. As a result, they risk either exposing themselves to permanent failure, or being taken to task for impacts they cannot realistically achieve.

Despite the conceptual simplicity of our model, we recognize that carrying it out poses significant challenges for managers as well as scholars. There is an urgent need for better knowledge on the challenges of scale and scope in the social sector. As noted, scale in the social sector can be achieved not only through

organizational growth, but also via myriad other means, particularly through influencing public policy, building collaborative networks, or the creation of new industries. However, even in terms of growth, the scaling rate is shockingly low: of over 200,000 new nonprofits created between 1975 and 2008, only 201 grew to revenues exceeding \$50 million (excluding hospitals and universities).³⁰ Much remains poorly understood. For each of the diverse means of achieving scale, what are the specific mechanisms and suitable industries? What are the main impediments to scale in terms of capital markets, capacity, and regulation—and how can they be addressed?³¹

Yet, there are many reasons for optimism. Social sector leaders and their funders are increasingly embracing performance measurement as critical to helping them achieve their missions at scale. They are shifting from a focus on evaluating impact *after* implementing their interventions, to also using measurement *during* program design and implementation in order to get real-time feedback for improving their work. This is true of large established organizations as well as new entrants to the field. For example, international NGOs such as CARE, Oxfam, and ActionAid International are all building monitoring and learning systems that can better enable mid-course correction. Large philanthropic foundations such as Gates, Hewlett, Rockefeller, and Edna McConnell Clark have all built up strong evaluation departments as part of a deeper commitment to outcome-oriented philanthropy.³² At the same time, a new generation of social entrepreneurs and impact investors are all demanding better data for informing decision making; they are supported in this aim by infrastructure-builders like the Global Impact Investing Network that are developing standardized metrics for use by investors. In terms of performance measurement methodologies, there has been a surge in the development of more participatory and integrative tools such as constituency feedback, most significant changes techniques, and developmental evaluation, as well as network-based approaches such as collective impact and outcome mapping (see Appendix 1)³³ suited to settings of high complexity involving interactions across multiple organizations and sectors.

In other words, the social sector is in a period of vibrant innovation on performance measurement. Given this moment of ferment, it is useful to step back and gain some clarity regarding the unit of analysis. The social sector has had a long history of program evaluation, mainly because funding has traditionally been directed to programs rather than to the organizations that design and deliver them. Organizations delivering multiple programs have often found that programmatic success does not equate to success in fulfilling their overall organizational missions. For that to happen, the various programs have to hold together in an integrated fashion. Our framework focuses squarely on the operational mission of the organization, under which multiple programs can be strategically combined.

There is a unique opportunity for funders to integrate multiple levels of analysis —programmatic, organizational, and societal—in assessing and improving performance in the social sector. This can be done in at least two ways. First, funders should allocate greater resources to building the management capacity of nonprofits and social enterprises to enable them to better integrate their own

programs. Funders should view themselves as part of a syndicate to enable mission success, even though a large part of their funding may be earmarked for programs. The three examples cited—Acumen, Robin Hood Foundation, and Millennium Challenge Corporation—all devote resources to such **organizational capacity building**, but they are the exception rather than the rule. The general trend has been in the opposite direction: towards a reduction in capacity building and other overhead costs in favor of direct funding to programs where the results are easier to measure. **However, the experience of venture capital, and increasingly of venture philanthropy, has shown that investments in human and systems capacities are foundational to generating longer-term results.** It is difficult for an operating organization to think strategically when it is at the mercy of different funders with different programmatic goals.

Second, funders should turn their attention to their own performance and impact, while easing off on their demands for operating organizations to prove their impacts. Foundations, governments, and impact investors are uniquely positioned—at a higher level in the organizational ecosystem than their grantees or investees—to see how the work of several operating organizations fits together to achieve impacts that are greater than the sum of the parts. They are also better resourced than most nonprofits and social enterprises to commission such research. Impacts on systemic societal problems are unlikely to be achieved by organizations acting alone; it thus makes more sense for funders rather than operating organizations to take on the challenge of measuring those impacts. In other words, nonprofits and social enterprises that operate in a niche should measure their outputs, and sometimes their outcomes, within that niche. Funders that operate higher up in the ecosystem should measure impacts at a societal level. While a handful of foundations (such as Gates, Hewlett, and McConnell Clark) are already oriented in that direction, the vast majority of funders, both institutional and individual, are still in the business of providing support for programs yet expecting systemic benefits. Bridging these multiple gaps—between performance at the level of programs, organizations, and society—will require funders to think more strategically about the different organizations and programs they fund so that their collective investment can achieve systemic impacts, with each individual piece playing its role. As with any organization in the social sector, doing so will require clarity about their operational mission, scale, and scope.

APPENDIX I. Methods for Assessing Social Performance (continued on next page)

Method	Key Advances	Innovators and Users (examples)	Sample Resources
Logic Models	<ul style="list-style-type: none"> ▪ Identifies and links the key results expected from an intervention: inputs, activities, outputs, outcomes and impacts. 	<ul style="list-style-type: none"> ▪ United States (USAID) ▪ European Commission (EC) ▪ W.K. Kellogg Foundation ▪ Innovation Network ▪ Urban Institute ▪ United Way of America 	<ul style="list-style-type: none"> ▪ http://usaidsite.carana.com/content/logical-framework-1f ▪ http://pdf.usaid.gov/pdf_docs/PNABN963.pdf ▪ http://ec.europa.eu/europeaid/multimedia/publications/documents/tools/europeaid_adm_pcm_guidelines_2004_en.pdf ▪ http://www.wkkf.org/knowledge-center/resources/2006/02/wk-kellogg-foundation-logic-model-development-guide.aspx ▪ http://www.innonet.org/client_docs/File/Logic_model_workbook.pdf ▪ http://www.urban.org/publications/310776.html ▪ http://www.younitedway.org/outcome-measurements ▪ http://www.thesroinetwork.org ▪ http://www.redf.org/learn-from-redf/publications/119 ▪ http://acumen.org/blog/news/the-method-behind-our-metrics/ ▪ http://acumen.org/wp-content/uploads/2007/01/BACO%20Concept%20Paper_01.24.071.pdf ▪ http://www.robinhood.org/metrics ▪ http://www.mcc.gov/pages/activities/activity/economic-rates-of-return
Expected Return and Cost Effectiveness	<ul style="list-style-type: none"> ▪ Measures social value of investments by quantifying benefits and costs, thus enabling comparison across potential investments. Used primarily by investors to optimize social value per unit of investment, often referred to as impact investing 	<ul style="list-style-type: none"> ▪ Social Return on Investment (SROI) Network ▪ REDF – social return on investment ▪ Acumen Fund – best available charitable option (BACO) ▪ Robin Hood Foundation – benefit-cost ratios (BCR) ▪ Millennium Challenge Corporation – economic rate of return (ERR) ▪ Abdul Latif Jameel Poverty Action Lab, MIT ▪ Innovations for Poverty Action 	
Experimental Methods	<ul style="list-style-type: none"> ▪ Assesses the impact of an intervention, and measures the size of that impact, by comparing outcomes in groups that receive an intervention to those that do not. 		<ul style="list-style-type: none"> ▪ http://www.povertyactionlab.org/methodology ▪ http://poverty-action.org/provenimpact

APPENDIX I. Initial Data Sources (continued from previous page, continued on next page)

Method	Key Advances	Innovators and Users (examples)	Sample Resources
Systematic Reviews	<ul style="list-style-type: none"> Sums up the best available research on an issue or question through synthesizing results of numerous studies. 	<ul style="list-style-type: none"> Campbell Collaboration International Initiative for Impact Evaluation (3ie) 	<ul style="list-style-type: none"> http://www.campbellcollaboration.org/ www.3ieimpact.org
Participatory and Relationship-Based Methods	<ul style="list-style-type: none"> Constituency Feedback and Perception Reports Most Significant Changes; Story-Based Evaluation Participatory Rural Appraisal (PRA) and variants Participatory Assessment Poverty <p>Solicits the perceptions of constituents about performance (e.g., grantee feedback on funders, community feedback on NGOs) that is then benchmarked against peers.</p> <p>Identifies impacts and changes without the use of preset indicators, relying on field-level stories of change and the systematic selection of the most significant of these stories by designated panels.</p> <p>Draws out knowledge of local people in analyzing their own realities as a basis for planning anti-poverty interventions. Outsiders serve as facilitators and listeners rather than experts.</p> <p>Adapts PRA techniques to policy issues, thus including the poor in discussions on policy priorities.</p>	<ul style="list-style-type: none"> Center for Effective Philanthropy Keystone Accountability Youth Truth M&E News (Rick Davies) Clear Horizon (Jess Dart) Institute of Development Studies (Robert Chambers) International Institute for Environment & Development U.N. Food and Agriculture Organization (FAO) World Bank 	<ul style="list-style-type: none"> http://www.effectivephilanthropy.org/index.php?page=assessment-tools http://www.keystoneaccountability.org/about/publicreport/what http://www.youthtruthsurvey.org/students http://www.mande.co.uk/docs/MSCGuide.pdf http://clearhorizon.com.au/category/resource-hub/most-significant-change/ http://www.ids.ac.uk/publication/from-pra-to-pla-to-pluralism-practice-and-theory http://www.iied.org/participatory-learning-action http://www.fao.org/docrep/W3241E/w3241e09.htm http://go.worldbank.org/QAASG4TK80

APPENDIX I. Initial Data Sources (continued from previous page)

Method	Key Advances	Innovators and Users (examples)	Sample Resources
<p>Integrative Approaches</p> <ul style="list-style-type: none"> ▪ Strategic Learning, Planning, and Evaluation ▪ Collective Impact, Community Change, and Complex Systems 	<ul style="list-style-type: none"> ▪ Links performance measurement to strategy using methods such as formative and developmental evaluation, balanced scorecards, strategy maps, dashboards and related tools. ▪ Brings together organizations across sectors to solve social problems by building a common agenda and shared measures of success. ▪ Emergent set of ideas and approaches for understanding complex, nonlinear, and adaptive systems that cannot be precisely predicted. 	<ul style="list-style-type: none"> ▪ Bridgespan Group ▪ FSG Social Impact Advisors ▪ McKinsey & Company Social Sector Office ▪ Ascendant Strategy Management Group ▪ Blue Avocado and CompassPoint ▪ ActionAid International ▪ iScale (for networks) ▪ Grassroots Business Fund ▪ FSG (collective impact) ▪ International Development Research Centre (outcome mapping) ▪ Full Frame Initiative ▪ New England Complex Systems Institute 	<ul style="list-style-type: none"> ▪ http://www.bridgespan.org/Publications-and-Tools/Performance-Measurement.aspx ▪ http://www.fsg.org/OurApproach/StrategicEvaluation.aspx ▪ http://lsi.mckinsey.com/ ▪ http://www.ascendantsmg.com/wiki/index.cfm/Balanced_Scorecard ▪ http://www.blueavocado.org/content/nonprofit-dashboard-and-signal-light-boards ▪ http://www.actionaidusa.org/publications/alps-accountability-learning-and-planning-system ▪ http://www.scalingimpact.net/files/IDRC_Network_IPARL_Paper_Final_0.pdf ▪ http://www.gbfund.org/sites/default/files/IPAL%20Framework%20Factsheet.pdf ▪ http://www.fsg.org/OurApproach/CollectiveImpact.aspx ▪ http://www.outcomemapping.ca/ ▪ http://fullframeinitiative.org/resources/evaluation/ ▪ http://www.necsi.edu/research/international/

Notes

1. See, for example, A. Ebrahim and E. Weisband, eds., *Global Accountabilities: Participation, Pluralism, and Public Ethics* (Cambridge: Cambridge University Press, 2007); L. Jordan and P. Van Tuijl, eds., *NGO Accountability: Politics, Principles and Innovations* (London and Sterling, VA: Earthscan, 2006); K.P. Kearns, *Managing for Accountability: Preserving the Public Trust in Nonprofit Organizations* (San Francisco, CA: Jossey-Bass, 1996).
2. See, for example, P. Brest and H. Harvey, *Money Well Spent: A Strategic Plan for Smart Philanthropy* (New York, NY: Bloomberg Press, 2008); L.R. Crutchfield and H.M. Grant, *Forces for Good: The Six Practices of High-Impact Nonprofits* (San Francisco, CA: Jossey-Bass, 2008); Monitor Institute, "Investing for Social and Environmental Impact: A Design for Catalyzing an Emerging Industry," Monitor Institute, 2009.
3. H. Hwang and W.W. Powell, "The Rationalization of Charity: The Influences of Professionalism in the Nonprofit Sector," *Administrative Science Quarterly* 54/2 (June 2009): 268-298; M. Power, *The Audit Society: Rituals of Verification* (Oxford: Oxford University Press, 1999).
4. P. Brest, H. Harvey, and K. Low, "Calculated Impact," *Stanford Social Innovation Review*, 7/1 (Winter 2009) 50-56; M. Marino, *Leap of Reason: Managing to Outcomes in an Era of Scarcity* (Washington, D.C.: Venture Philanthropy Partners, 2011); T.J. Tierney and J.L. Fleishman, *Give Smart: Philanthropy that Gets Results* (New York, NY: Public Affairs, 2011).
5. C. Roche, *Impact Assessment for Development Agencies: Learning to Value Change* (Oxford: Oxfam GB, 1999), p. 21.
6. Crutchfield and Grant, op. cit., p. 24.
7. H. Jones, "The 'gold standard' is not a silver bullet for evaluation," London, Overseas Development Institute, 2009.
8. L. Bickman, "The Functions of Program Theory," in L. Bickman, ed., *Using Program Theory in Evaluation, New Directions for Program Evaluation* (San Francisco, CA: Jossey-Bass, 1987); P.J. Rogers, "Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions," *Evaluation*, 14/1 (January 2008): 29-48; C.H. Weiss, *Evaluation Research: Methods for Assessing Program Effectiveness* (Englewood Cliffs, NJ: Prentice-Hall, 1972); J. Wholey, *Evaluation: Promise and Performance* (Washington, D.C.: Urban Institute Press, 1979).
9. J. C. Sawhill and D. Williamson, "Mission Impossible? Measuring Success in Nonprofit Organizations," *Nonprofit Management and Leadership* 11/3 (Spring 2001): 371-386.
10. United Way of America, "Agency Experiences with Outcome Measurement: Survey Findings," Report 0196, Alexandria, VA, 2000.
11. P.D. Hall, "A Solution is a Product in Search of a Problem: A History of Foundations and Evaluation Research," 2005, <www.hks.harvard.edu/fs/phall/EVALUATION.pdf>: 33.
12. P. Frumkin, *Strategic Giving: The Art and Science of Philanthropy* (Chicago, IL: University of Chicago, 2006).
13. M. Kramer, R. Graves, J. Hirschhorn, and L. Fiske, "From Insight to Action: New Directions in Foundation Evaluation," FSG Social Impact Advisors, 2007, p. 5.
14. "Interview—Paul Brest, Jed Emerson, Katherina Rosqueta, Brian Trelstad and Michael Weinstein," *Alliance Magazine*, April 1, 2009, <www.alliancemagazine.org/node/2128>; B. Trelstad, "Simple Measures for Social Enterprise," *Innovations*, 3/3 (2008); M.M. Weinstein (with C.E. Lamy), "Measuring Success: How Robin Hood Estimates the Impact of Grants," New York, NY, Robin Hood Foundation, 2009.
15. B. Glasrud, "The Muddle of Outcome Measurement: Be Careful How You Measure Programs," *Nonprofit World*, 19/6 (November/December 2001): 35; S. Torjman, "Are Outcomes the Best Outcome?" Ottawa, Caledon Institute of Social Policy, 1999.
16. G. Hernández and M.G. Visser, "Creating a Culture of Inquiry: Changing Methods—and Minds—on the Use of Evaluation in Nonprofit Organizations," San Francisco, CA, The James Irvine Foundation, 2001, p. 2.
17. The challenges of performance measurement, scale, and scope facing these two organizations are elaborated in two cases: "Aravind Eye Hospital, Madurai, India: In Service for Sight," Harvard Business School Case Number 593098, 2009; "Harlem Children's Zone: Driving Performance with Measurement and Evaluation," Harvard Business School Case Number 303109, 2004.
18. T. Vickers, "Driving Down the Cost of High-Quality Care: Lessons from the Aravind Eye Care System," interview with Dr. Srinivasan, *Health International*, 11 (2011): 23.
19. For resources on theory of change, see: ActKnowledge, "Theory of Change," 2010, <www.theoryofchange.org>; Andrea A. Anderson, "Theory of Change as a Tool for Strategic Planning: A Report on Early Experiences," New York, NY, Aspen Institute Roundtable on Community

- Change, 2004, <www.aspeninstitute.org/policy-work/community-change/publications>; Keystone Accountability for Social Change, "Developing a Theory of Change," London, 2008, <www.keystoneaccountability.org/about/theoryofchange>.
20. Harlem Children's Zone, "2010-2011 Biennial Report," New York, 2011, accessed August 22, 2013, <http://hcz.org/books/HCZ_2011_Biennial/index.html>.
 21. V.K. Rangan, "Lofty Missions, Down-to-Earth Plans," *Harvard Business Review*, 82/3 (March 2004): 112-119.
 22. R.M. Kanter, "Even Bigger Change: A Framework for Getting Started at Changing the World," Harvard Business School Background Note 305099, 2005; R.M. Kanter, "The Change Wheel: Elements of Systemic Change and How to Get Change Rolling," Harvard Business School Background Note 312-083, 2011; P. Uvin, P. Jain, and L.D. Brown, "Think Large and Act Small: Toward a New Paradigm for NGO Scaling Up," *World Development*, 28/8 (August 2000): 1409; P. Uvin and D. Miller, "Paths to Scaling-up: Alternative Strategies for Local Non-governmental Organizations," *Human Organization*, 55/3 (Fall 1996): 344.
 23. World Microfinance Forum Geneva, "Is it fair to do business with the poor? Report on the debate between Muhammad Yunus and Michael Chu," Geneva, 2008.
 24. P. Brest, "A Decade of Outcome-Oriented Philanthropy," *Stanford Social Innovation Review*, 10/2 (Spring 2012): 42-47; Frumkin, op. cit.
 25. A. Ebrahim, *NGOs and Organizational Change: Discourse, Reporting and Learning* (Cambridge: Cambridge University Press, 2003); R. Eyben, "Uncovering the Politics of 'Evidence' and 'Results': A Framing Paper for Development Practitioners," April 2013, <<http://bigpushforward.net/wp-content/uploads/2011/01/Uncovering-the-Politics-of-Evidence-and-Results-by-Rosa-lind-Eyben.pdf>>; T. Wallace, L. Bornstein, and J. Chapman, *The Aid Chain: Coercion and Commitment in Development NGOs* (Rugby, UK: Intermediate Technology Publications, 2006).
 26. The measurement challenges facing these three organizations are elaborated in three cases developed by the authors: A. Ebrahim and V.K. Rangan, "The Millennium Challenge Corporation and Ghana," Harvard Business School Case Number 310-025, 2010; A. Ebrahim and V.K. Rangan, "Acumen Fund: Measurement in Venture Philanthropy (A)," Harvard Business School Case Number 310-011, 2011; A. Ebrahim and C. Ross, "The Robin Hood Foundation," Harvard Business School Case Number 310-031, 2010.
 27. Trelstad, op. cit.
 28. S. Earl, F. Carden, and T. Smutylo, *Outcome Mapping: Building Learning and Reflection into Development*; J. Kania and M. Kramer, "Collective Impact," *Stanford Social Innovation Review*, 9/1 (Winter 2011): 36-41.
 29. V.K. Rangan, "African Comprehensive HIV/AIDS Partnerships (ACHAP): The Merck/Gates Initiative in Botswana," Harvard Business School Case Number 505-057, 2007.
 30. P. Kim and J. Bradach, "Why More Nonprofits are Getting Bigger," *Stanford Social Innovation Review*, 10/2 (Spring 2012): 15-16.
 31. R.S. Kaplan and A.S. Grossman, "The Emerging Capital Market for Nonprofits," *Harvard Business Review*, 88/10 (October 2010): 110-118.
 32. Brest (2012), op. cit.
 33. D. Bonbright, D. Campbell, and L. Nguyen, "The 21st Century Potential of Constituency Voice: Opportunities for Reform in the United States Human Services Sector," Alliance for Children & Families, United Neighborhood Centers of America, and Keystone Accountability, 2009; R. Davies and J. Dart, "The 'Most Significant Change' (MSC) Technique: A Guide to Its Use," (2005), <www.mande.co.uk/docs/MSCGuide.pdf>; Earl, Carden, and Smutylo, op. cit.; S. Khagram, C. Thomas, C. Lucero, and S. Mathes, "Evidence for Development Effectiveness," *Journal of Development Effectiveness*, 1/3 (2009): 247-270; F. Twersky, P. Buchanan, and V. Threlfall, "Listening to Those Who Matter Most, the Beneficiaries," *Stanford Social Innovation Review*, 11/2 (Spring 2013): 41-45; M.Q. Patton, *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use* (New York, NY: Guilford Press, 2011).