

Fletcher Executive Education Program Reading Material

Help us keep in compliance with copyright law.

We are providing this scan of a required reading for enrolled students in this course. Please do not print, download, or otherwise share the link to this material with anyone outside of this course. To do otherwise would be a likely violation of copyright law.

Thank you for your understanding.

MEASURING SOCIAL CHANGE

This page intentionally left blank

MEASURING SOCIAL CHANGE

Performance and Accountability in
a Complex World

Alnoor Ebrahim

STANFORD BUSINESS BOOKS, AN IMPRINT OF STANFORD UNIVERSITY PRESS

Stanford, California

Stanford University Press

Stanford, California

© 2019 by the Board of Trustees of the Leland Stanford Junior University.

All rights reserved.

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or in any information storage or retrieval system without the prior written permission of Stanford University Press.

Special discounts for bulk quantities of Stanford Business Books are available to corporations, professional associations, and other organizations. For details and discount information, contact the special sales department of Stanford University Press. Tel: (650) 725-0820, Fax: (650) 725-3457

Printed in the United States of America on acid-free, archival-quality paper

Library of Congress Cataloging-in-Publication Data

Names: Ebrahim, Alnoor, author.

Title: Measuring social change : performance and accountability in a complex world / Alnoor Ebrahim.

Description: Stanford, California : Stanford University Press, 2019. | Includes bibliographical references and index.

Identifiers: LCCN 2018060866 (print) | LCCN 2019001436 (ebook) | ISBN 9781503609211 (e-book) | ISBN 9781503601406 (cloth : alk. paper)

Subjects: LCSH: Nonprofit organizations—Evaluation. | Social entrepreneurship—Evaluation. | Organizational effectiveness—Measurement. | Social change—Evaluation.

Classification: LCC HD62.6 (ebook) | LCC HD62.6 .E277 2019 (print) | DDC 361.7068/4—dc23

LC record available at <https://lcn.loc.gov/2018060866>

Cover design: Michel Vrana

Cover illustration: Plate XIV–1 from *Interaction of Color*, 50th Anniversary Edition, by Josef Albers, courtesy of Yale University Press.

Typeset by Newgen in 11.25/15 Baskerville

for Shirin and Sadrudin
beyond measure

This page intentionally left blank

CONTENTS

Acknowledgments	ix
Introduction	1
1. Conceptualizing Social Performance in a Complex World	9
2. Niche Strategy	51
3. Integrated Strategy	84
4. Emergent Strategy	122
5. Ecosystem Strategy	160
6. Designing Social Performance Systems	192
7. Roles of Funders	208
Appendix: Practical Resources	245
Notes	251
References	265
Index	291

INTRODUCTION

THE IDEA FOR THIS BOOK crystallized on an icy winter day in Ottawa. I was attending a meeting with representatives of the bilateral aid agencies from six countries—Canada, Denmark, The Netherlands, Norway, Sweden, and the United Kingdom—all of whom were under pressure from their governments to demonstrate value for taxpayer money allocated to international development. Collectively, they had funded hundreds, if not thousands, of civil society organizations around the world on social issues as diverse as poverty alleviation, the environment, education, health care, and agricultural development.

I had been invited to suggest a methodology for assessing how much social benefit their investments had produced. But as the discussions with these senior managers progressed, it became increasingly clear that existing tools for measuring social performance fell far short of addressing their complex needs and constraints. It was at this meeting that I first probed a new approach to measurement that would eventually become this book.

Leaders across the social sector—in nonprofit and nongovernmental organizations, social enterprises, international aid agencies, philanthropic foundations, and impact investors—have long struggled with how to measure social change. Unlike assessing the performance of a business, there is no common currency of measurement. Money is

typically a means rather than an end. Results are not easily comparable across organizations with diverse social goals such as fighting poverty, improving public health, reducing climate change, or advancing human rights. And time horizons are long, sometimes taking years or even a generation to bear results.

The significance of this challenge has only grown as the sector has burgeoned in revenue and size. In the United States alone, there are over 1.5 million registered nonprofit organizations accounting for over \$2 trillion in annual revenues.¹ In addition, the rapidly emerging global impact investing industry represents a market of several hundred billion dollars.² And for a new generation of donors and investors, allocating money to a worthy cause is no longer sufficient—measurable results are essential.

Over the past decade, I have had the good fortune of engaging with hundreds of leaders in the sector on these challenges through executive education at the Harvard Business School, Harvard Kennedy School, and the Fletcher School of Law and Diplomacy. I have also worked closely with practitioners on a working group established by the G8 to create global guidelines on measuring social impact, and through advisory roles with impact investors and industry associations such as the Global Impact Investing Network. These experiences have shaped my thinking and the research in this book.

In the pages that follow, I tackle three main challenges of performance common across the social sector: *what to measure*, what kinds of *performance measurement and management systems* to build, and how to align multiple demands for *accountability*. Over the past ten years, the sector has become consumed by the mantras of “impact” and “accountability”—an expectation that organizations must be held to account for their performance in solving difficult social problems. A growing body of literature has emerged to help managers to clarify their models of cause and effect (theory of change), to develop indicators along that causal path (logic models), and to conduct rigorous assessments of their interventions (impact evaluation). This work has no doubt advanced the field.

But the push for impact is also underpinned by unquestioned assumptions: that long-term outcomes are more valuable than short-term outputs, that clarity about cause and effect is a precondition to good strategy, and that everyone can agree on what constitutes good performance. On the face of it, these are rational expectations about performance fueled by an increasingly competitive funding environment. After all, who could object to payment based on solid evidence of performance? Yet, in a world of messy problems that lack straightforward solutions, there remains a substantial gap between expectations and reality.

Unambiguous evidence of cause and effect turns out to be rare in the arena of social change, and most managers have neither the expertise nor the resources to parse the social science research (about which scientists themselves often disagree). They are busy building their organizations in resource-constrained environments, and need information to help them with decisions today, not years down the road. Moreover, rarely can a single organization address a social problem by itself, as it typically lacks control over the various factors affecting an outcome. Yet, if it collaborates with others, it runs into the problem of how to isolate, measure, and claim credit for the impacts of its own work, especially when trying to convince potential funders to invest in it.

To unpack these challenges, I take a “contingency” approach to measurement—in which *what one measures* depends on a set of practical conditions or constraints experienced by the organization. What are the conditions under which it might make more sense to focus on short-term results, and others where the long-term is essential? What results can an organization reasonably measure, and legitimately take credit for, when it operates in an ecosystem of many other actors? These questions bear the mark of contingency. To address them, I untangle a pair of constraints widely experienced by social sector organizations: *uncertainty* about cause-effect knowledge for addressing a social problem and limited *control* over the activities and conditions necessary for producing long-term outcomes. Based on these contingencies,

I develop a framework for guiding managers in deciding what they should measure.

Building on this framing, I then identify the types of *performance measurement and management systems* (henceforth referred to as *performance systems*) best suited to different types of organizations. Put simply, different organizations need different types of performance systems. For instance, a provider of emergency medical services needs a performance system focused on quality control and standardization, whereas an advocacy organization needs a system that enables constant adaptation to changes in its policy environment. Drawing on a series of in-depth case studies, I identify four key types of performance systems, and I provide a basis for managers to decide what type of system might best fit their needs.

This book also unpacks a third closely related challenge: the dilemma of *multiple accountabilities*. Even when organizations have reliable measures of their social performance, those measures are unlikely to satisfy key stakeholders when their interests are divergent or in conflict. Such stakeholders may include various funders, regulators, and even different types of clients or beneficiaries. Social sector leaders often speak with exasperation about the many demands for accountability they face, as though caught in a web, struggling in many directions but getting nowhere. Their key task, however, is not to be more accountable to everyone for everything but rather to align and prioritize among competing accountability demands. The structural problem is that “upward” accountability to funders tends to be well developed for the simple reason that funders control resources, whereas means of “downward” accountability that give voice and influence to clients or beneficiaries are typically weak. It thus falls upon social sector leaders to be clear about their accountability priorities and to design their performance systems accordingly. They are not simply on the receiving end of accountability, but have agency in shaping its terms.

These broad themes at the heart of this book—what to measure, what kinds of performance systems to build, and how to set the terms of accountability—point to a couple of fundamentally different views

on the purpose of performance measurement. The first is to enable better organizational decision making. In serving this purpose, measurement is seen as a rational-technical process that develops objective metrics and procedures for achieving organizational goals. It enables midcourse corrections (monitoring) while also assessing progress (evaluation), through instruments such as logic models, scorecards, and randomized control trials. This function of measurement appeals to the agent of change in each of us who wishes to use information to make better decisions and, consequently, to deliver better results for society. This approach to performance measurement is consistent with the predominant literature on strategy and outcome management, and it is the primary perspective adopted in this book.

But at the same time, as anthropologists and sociologists have shown, measurement is hardly a purely rational endeavor. It is an act of social construction, where metrics are imbued with the values of actors, signaling what is important and what is not.³ The challenge of multiple accountabilities arises from this aspect of measurement, where stakeholders apply varying yardsticks to judging the performance of an organization, often perceiving its worth differently. It is thus not uncommon for organizations to tailor their metrics and communications to their stakeholders in an effort to secure not just funding but social legitimacy. Accountability, as such, is a relationship of power; it does not stand objectively apart from the expectations and demands of external actors. Social sector leaders must understand both of these roles of measurement, rational and social, in building their performance systems.

Structure of this Book

The first chapter provides the conceptual foundation for the book, introducing a framework based on the two contingent factors that managers must understand in developing their performance systems: uncertainty about cause-effect, and control over outcomes. These two factors provide a basis for differentiating among four types of social

change strategies: niche, integrated, emergent, and ecosystem strategies. The key managerial implication is that each type of strategy requires a distinct type of performance system.

The ensuing four chapters use case studies to illustrate and develop each of these strategy types and their attendant performance systems. Chapter 2 examines a *niche strategy* in an emergency medical response service, where the relationship between cause and effect is relatively well understood, but there is low control over long-term outcomes. This strategy requires a performance system based on *standardization* to produce outputs of reliable quality. The broader value of this case lies in its illumination of the key features of a performance system needed by any organization with a niche strategy. Following a similar logic, Chapter 3 considers an *integrated strategy* in rural agricultural development, where the relationships between cause and effect are relatively well understood but where multiple interventions must be combined in order to produce long-term outcomes. This strategy requires a performance system based on *coordination* to produce, prioritize, and sequence outputs that generate interdependent outcomes.

Chapters 4 and 5 move on to organizational contexts where relationships about cause and effect are poorly understood or complex, typically involving many moving parts and players. The first of these chapters examines an *emergent strategy* in global policy advocacy work, where an organization's ability to control policy outcomes is severely constrained. This strategy requires a performance system based on *adaptation* that allows the organization to recognize and take advantage of new opportunities to influence key actors within its system. Chapter 5 turns to an *ecosystem strategy* in addressing urban homelessness where, although cause-effect relationships are complex, it is possible to increase control over outcomes by aligning the work of many interdependent actors. This strategy requires a performance system based on *orchestration* that restructures relationships among key players in a system in order to generate joint outcomes.

Chapter 6 brings together the key insights, synthesizing the learning from the case studies and revisiting key questions in the book:

What should an organization measure? What kinds of performance systems best fit its needs? What should be its accountability priorities? I provide a side-by-side comparison of the performance systems used in each of the cases, identifying not only what they have in common, but also how they differ in fundamental ways. This analysis leads to a typology of distinct performance systems, each suited to supporting one of the four strategies noted above.

Finally, in Chapter 7, I turn to the roles of funders in supporting performance measurement and management. Funders, be they grant makers or investors, have the potential to help or hinder the organizations they support. Leaders of nonprofits and social enterprises often complain that their funders expect rigorous monitoring and evaluation of performance but are rarely willing to support it. There are exceptions, of course, and I draw on the experiences of three innovators—an impact investor, a grant-making foundation, and a bilateral aid agency—to illustrate the critical roles that funders can play in enabling better performance measurement, not only within their own portfolios but also in raising the bar among their peers. I distinguish between different stages of funder decision making—search, diligence, improvement, and evaluation—to show that each stage requires a different toolkit for performance measurement. The experiences of these pioneering funders also demonstrate how measurement can help to close the gap between upward accountability to funders and downward accountability to clients or beneficiaries.

This concluding chapter also highlights the measurement challenges that funders face, not only in comparing the performance of one grantee or investee to another, but also in assessing their aggregate performance as a portfolio. I argue that funders are much better positioned than their investees to achieving social impacts at a system level. Because they support and oversee many, sometimes hundreds, of nonprofits and social enterprises that typically act independently of one another, funders are poised to connect and leverage the combined impact of that work. Few funders have risen to this strategic

challenge, but they are uniquely situated in the organizational ecosystem to do so.

In a world beset by enormous social challenges, there is an urgent need to measure and improve the social performance of organizations, be they public, private, or nonprofit. Pressures on managers to be clear about results, and on investors to allocate capital based on social performance, will only increase. This book articulates a pluralistic rather than a singular model of performance measurement. I hope it will enable social sector leaders to be clear about what they aim to achieve, to be realistic about the constraints they face, and ultimately to have the courage to set their own terms of accountability.

Chapter 1

CONCEPTUALIZING SOCIAL PERFORMANCE IN A COMPLEX WORLD

We will need to give up childish fantasies that we can have total guarantees of others' performance. We will need to free professionals and the public service to serve the public. We will need to work towards more intelligent forms of accountability.

ONORA O'NEILL (2002)

IN A SERIES OF LECTURES on BBC radio, the Cambridge philosopher Onora O'Neill offered a provocative take on accountability in public service. She argued that efforts to improve the performance of public service providers, be they doctors or teachers or police officers, had a dark side: they were leading to a compliance-driven culture focused on rule-following behavior and quantitative targets rather than actually improving performance (O'Neill 2002). Her apprehensions can be extended to the social sector more broadly, including nonprofit organizations and social enterprises, where the mantras of "accountability" and "impact" have been ascendant for over a decade. Yet, despite the proliferation of reporting requirements, measurement procedures, and auditing rituals, there is limited evidence that performance in the sector has substantially improved (Ebrahim and Weisband 2007; Espeland and Stevens 2008; Hwang and Powell 2009; Lewis and Madon 2003; Power 1999).

What then might meaningful performance measurement based on "intelligent forms of accountability" look like? The purpose of this

chapter is to provide a way of thinking about performance and accountability that is strategy driven rather than compliance driven. I develop a pair of frameworks that enables social sector leaders to clarify what they realistically can and cannot achieve through their organizations, while simultaneously providing a basis for holding their own feet to the fire. In other words, the frameworks are devices that managers can use to specify their own terms of accountability.

This chapter is divided into two main parts. First, I provide a brief introduction to the foundations of organizational performance assessment, drawing from the literatures in business and nonprofit management, program evaluation, and development studies. I also take a closer look at the approaches to performance and impact measurement used in two practitioner communities: philanthropic foundations, impact investors, and nonprofit organizations (NPOs) based primarily in the United States; and organizations working in the field of international development such as bilateral aid agencies and non-governmental organizations (NGOs). These two communities are in the midst of starkly parallel dilemmas about impact measurement and accountability, although they operate almost independently. I identify the ongoing concerns about performance measurement in both communities, devoting special attention to the uses and limitations of “logic models” that have been foundational to both.

In the second half of the chapter, I build on this analysis to develop two frameworks for measuring and improving social performance. The first is a general model of social sector performance comprising three core components: an organization’s value proposition, its model of social change, and its accountability priorities. All organizations need to be clear about these components if they are to make systematic and measurable progress in addressing social problems.

I then build on this general model to develop a more nuanced “contingency framework” for social performance. I argue that what an organization should measure, and consequently should be held accountable for, depends on two key factors that vary from organization to organization: *uncertainty* about cause and effect, and *control*

over outcomes. The framework offers a strategic basis for deciding what to measure, while recognizing the difficult constraints facing managers in impacting social problems. This contingency approach suggests that—given the varied work, aims, and environments of social sector organizations—some organizations should be measuring long-term outcomes while others would be better off focusing on short-term outputs. More importantly, I offer a logic for determining which kinds of measures are appropriate, given not only the organization's goals but also its position within a larger ecosystem of actors.

My normative argument, embedded in this contingency framework, is that it is not feasible or even desirable for all organizations to develop metrics that run the full gamut from outputs to societal outcomes. The more important challenge is one of aligning measurement with goals and strategy, especially the goals that an organization can reasonably control or influence. I contend that organizational efforts extending beyond this scope are a misallocation of scarce resources. For many social sector leaders, there is a temptation to overreach, to claim credit for social changes that may be beyond their actual control, in order to secure funding and social legitimacy. The challenge for managers is to be more realistic and grounded in framing the performance of their organizations, and thus to better achieve goals within their control.

Some readers will no doubt be troubled by my argument that not all social sector organizations should be measuring long-term outcomes and impacts. After all, if they don't measure outcomes, how will we ever know if they are making a difference? This reasoning fails to recognize that social change is contingent on many factors—that organizations vary in their goals, their knowledge about cause and effect, and in their interdependence with other actors in their ecosystems. The purpose of a contingency framing of social performance is to unpack these differentiating features, so that managers can be realistic about what they aim to achieve and then measure and improve performance accordingly.

Conceptual Foundations

Much of the current writing on the performance of organizations is rooted in the vast literature on organizational effectiveness, which has long identified three basic types of indicators for judging organizational performance: outcomes, processes, and structures (Goodman and Pennings 1977; Scott 1977; Suchman 1967). Outcomes are forward-looking measures in that they are the results predicted from a set of outputs such as goods or services; processes are measures of effort that focus on inputs and activities carried out by organizations; and structural indicators assess the capacity of an organization to perform work. Of these three types of indicators, organizational sociologists have noted that outcomes are often considered “the quintessential indicators of effectiveness, but they also may present serious problems of interpretation” such as inadequate knowledge of cause and effect, the time periods required to observe results, and environmental characteristics beyond the control of the organization such as market conditions or receptivity of external stakeholders (Scott 1992, 354).

The vast literature on organizational performance and effectiveness appears to converge on one key insight: there are rarely any singular and unambiguous measures of success in organizations. Even in for-profit firms, where it is tempting to assume that outcome metrics are unambiguous because of the profit motive, this turns out rarely to be the case. Meyer (2002; Meyer and Gupta 1994) identifies four broad types of measures common in profit-making businesses: valuation of the firm in capital markets, accounting measures, nonfinancial measures, and cost measures. Not only is there no single measure that is adequate for measuring firm performance, but some metrics can even point in opposite directions.

For example, key accounting measures (such as return on investment, return on assets, cash flows, and other measures of sales and profit) are not necessarily correlated with market measures (such as market value, return on equity, and change in share prices). It is not uncommon for a firm that fails to turn a profit to nonetheless see an uptick in its share price, or for a firm that makes considerable short-

term profit to lose the confidence of long-term investors. In short, firms tend to use multiple performance measures simultaneously, with the value of these different measures resting in the fact that they don't correlate with one another—a characteristic that Meyer (2002; Meyer and Gupta 1994) has called the “performance paradox”—for if the measures did correlate, it would be possible to rely on a single roll-up metric. The main point here is that even in a sector where there is general convergence around profit, there is a need for multiple simultaneous measures in order to judge performance.

These challenges are even more pronounced and complex in the social sector (Ebrahim and Rangan 2014; Stone and Cutcher-Gershenfeld 2001). Financial measures are generally treated as an input rather than an outcome, and there is wide variation across industries on what constitutes a valuable outcome (Anthony and Young 2004; Oster 1995). Nonprofit ratings agencies that have traditionally relied on efficiency ratios such as program-to-administration expenses for rating performance are now widely criticized even by their advocates for being too narrow and misleading (Philanthropy Action 2009). A primary metrics challenge remains in establishing reliable and comparable nonfinancial measures. While there appears to be a growing convergence around the notion of “impact” as the ultimate nonfinancial measure of performance, there remains considerable ambiguity around how to operationalize it and whether it helps or hinders organizations in managing performance.

Moreover, because ownership is generally less clear in nonprofits and hybrid social enterprises than it is in for-profit firms, this can lead to demands for accountability and reporting from multiple funders (such as foundations, private investors, government agencies, and individual donors) and varying expectations about performance from clients, communities, regulators, taxpayers, and their own staff and boards (Edwards and Hulme 1996b; Kearns 1996; Lindenberg and Bryant 2001; Najam 1996a; Oster 1995). Some scholars have suggested that there are as many types of accountability as there are distinct relationships among people and organizations (Lerner and Tetlock 1999),

characterizing the pronounced nature of this condition in social sector organizations as “multiple accountabilities disorder” (Koppell 2005).

Despite these many challenges, there have been important advances in the social sciences on the measurement of social performance. To anchor our discussion of these developments, I draw on a long tradition of research in program evaluation that offers a body of theory and methods for the design and assessment of social programs (e.g., Bickman 1987; McLaughlin and Jordan 1999; Rogers 2007; Weiss 1972). A foundational body of work in evaluation research is “program theory,” which provides a basis for conceptualizing, designing, and explicating social programs; for understanding the causal linkages (if-then relationships) between program processes and outcomes; and for diagnosing the causes of trouble or success (Blalock 1999; Funnell 1997; Rogers 2008; Rogers et al. 2000). Program theory may be seen as a method of applied social science research (Lindgren 2001) that allows for empirical testing of hypotheses embedded in any social program and thereby for advancing knowledge on the validity of program hypotheses in real-life environments (Chen 1990; Greene 1999; Weiss 1995).

A specific manifestation of program theory, the so-called logic model or “results chain,” has emerged as a dominant instrument through which organizations in the social sector identify their social performance metrics. Figure 1.1 depicts the key components of the logic model—inputs, activities, outputs, individual outcomes, and societal outcomes—including examples of the types of measures under each step. The direction of arrows in the figure, from left to right, emphasize the predictive, or propositional, aspect of the model and the measurement of performance as far down the chain as possible, in order not only to capture the causes (inputs and activities) and immediate goods or services delivered by an organization (outputs) but also to assess their long-term effects on the lives of individuals and communities or societies (outcomes).

The term *impact* has become part of the everyday lexicon of social sector funders in recent years, with frequent references to “high-

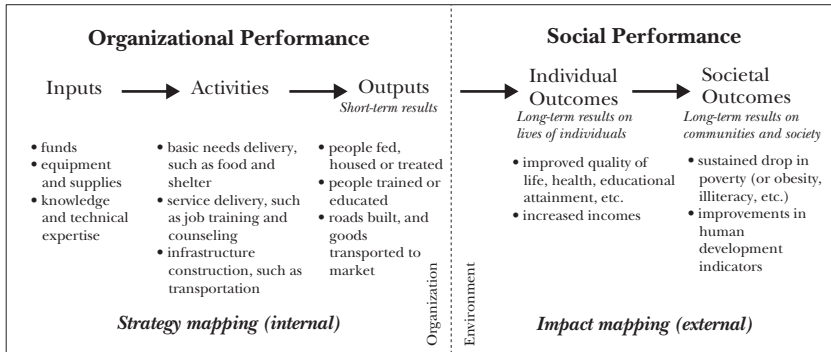


FIGURE 1.1 Logic Model and Results Mapping

impact nonprofits” or “impact philanthropy” and “impact on steroids” (Brest, Harvey, and Low 2009; Morino 2011; Tierney and Fleishman 2011). But the term 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” (Roche 1999, 21) or results that target the “root causes” of a social problem (Crutchfield and Grant 2008, 24). A widely used, if expansive, definition adopted by many international aid agencies explains impact as “the positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended” (Leeuw and Vaessen 2009, ix).

What most of these definitions share is an emphasis on causality—changes *brought about by* actions or effects *produced by* an intervention—suggesting that it is not sufficient to simply assess what happened after an intervention, but rather to assess whether those effects or changes can be causally linked to it (Brest and Harvey 2018; Jones 2009; White 2006). As such, an impact is the “difference made” by an intervention, be it short term or long term, and it may arise at individual, community, or societal levels. Many definitions, however, use the term *impact* to refer only to long-term societal changes. For example, a number of manuals on logic models describe impacts as occurring at the

level of organizations, communities, or systems after a period of many years (e.g., Knowlton and Phillips 2013, 38; W. K. Kellogg Foundation 2004, 2). In this book I opt for the former usage, reserving the term *impact* to mean the changes produced, or difference made, by an intervention or set of interventions. It is up to the organization to specify the nature of that impact—short term or long term, on individuals or society—and then to measure accordingly.¹

In order to identify their intended impacts, both short and long term, social sector organizations have increasingly turned to logic models and their variations such as logical framework analysis (LFA). The use of these instruments is often required by funders, and they have been diffused by a global industry of international development professionals, particularly professional evaluators employed by bilateral aid agencies and multilateral development banks (Roche 1999, 18–20), as well as by private philanthropic foundations seeking to measure the impacts of their grantmaking and to be more strategic about their giving (Brest 2012; Frumkin 2006; Morino 2011; Porter and Kramer 1999; Tierney and Fleishman 2011).

Although the logic model has emerged as a dominant instrument for clarifying metrics of social performance, using it as a planning tool is far from straightforward. Its utility is constrained by the complex and often poorly understood nature of cause-effect relationships for achieving social results. A major challenge in using the logic model for measuring performance is that it implicitly contains two causal chains, with the dividing line being at outputs or the organizational boundary, as represented by the dotted vertical line in Figure 1.1. On the left-hand side of the figure (inputs, activities, outputs), results are largely within the organization's control, and the causal logic is determined by strategic decisions on how to produce products or services. In elaborating this part of the results chain—organizational performance—the task facing social sector organizations is largely similar to that facing for-profit organizations. However, social sector organizations confront the additional challenge of establishing cause-effect relationships that occur *outside* their organizational boundaries—social performance—where organizational level activities and outputs (causes) are

expected to lead to outcomes on the lives of beneficiaries and society (effects).

The mapping of cause-effect relationships is, of course, also important to profit-seeking firms.² In particular, cause-effect relationships are integral to “balanced scorecards” that map causal relationships between the objectives and activities necessary for executing a strategy (Kaplan and Norton 1996, 2004). However, the cause-effect relations mapped by these concepts and tools are primarily internal to the organization (the left half of Figure 1.1) for establishing a pathway from activities to organizational-level results (Nørreklit 2000, 2003; Speckbacher, Bischof, and Pfeiffer 2003) rather than societal-level results. And although the concept of the balanced scorecard has been extended to social sector activities (Kaplan 2001), it has not been adapted to include the more complex cause-effect relations between outputs and outcomes that typically arise outside of organizational boundaries.³

In short, social sector organizations require attention to two cause-effect chains subsumed within the logic model: (1) a “strategy map” that links inputs and activities to outputs within the organization, comparable to those used by for-profit businesses; and (2) an “impact map” that links outputs to outcomes for assessing social performance. It is these complex causal logics of how organizational level results (outputs) transform into social change (outcomes) that are at the heart of the vexing challenges for performance measurement and accountability in the social sector.

Before elaborating a measurement framework that begins to address this challenge, it is useful to take stock of the current state of measurement practice. The experience of practitioner communities offers important insights for developing conceptual frameworks relevant to managerial decision making.

The Current State of Measurement Practice

Variations of the logic model and outcome management have become increasingly common in two practitioner communities that I discuss here: philanthropic foundations and impact investors seeking

to allocate their resources effectively, and international development organizations such as bilateral government agencies and nongovernmental organizations seeking to improve accountability and “value for money” in development aid. These two practitioner communities are at the epicenter of current debates and innovations in social performance. The academic literature in management, nonprofit studies, and philanthropy lags behind in providing theoretical and analytical insights to this growing field of practice. Because of the vastness and diversity of the social sector, I focus the discussion below on organizations working mostly in the field of poverty alleviation. This field is wide but well established, with an increasingly coherent set of debates on performance measurement.

Impact measurement debates have taken center stage in the world of American private philanthropy as foundations have placed renewed emphasis on measuring the impacts of their grantmaking and thinking more strategically about their giving (Frumkin 2006; Porter and Kramer 1999). This spirit of “smart philanthropy” was captured by Paul Brest and Hal Harvey during their tenure as presidents, respectively, of the Hewlett Foundation and the ClimateWorks Foundation:

[A]ccomplishing philanthropic goals requires having great clarity about what those goals are and specifying indicators of success before beginning a philanthropic project. . . . This, in turn, requires an empirical, evidence-based understanding of the external world in which the plan will operate. And it requires attending carefully to milestones to determine whether you are on the path to success with a keen eye for signals that call for midcourse corrections. These factors are the necessary parts of what we regard as the essential core of strategic philanthropy—the concern with *impact*. (Brest and Harvey 2008, 7, emphasis in the original)

This concern with impact has fueled efforts to quantify it in order to allocate funding among competing possibilities—what Brest and Harvey have called “impact on steroids.” And it has attracted considerable attention in the burgeoning field of impact investing, charac-

terized by investors that intentionally target social and financial objectives and seek to measure the achievement of both (OECD 2015; Social Impact Investment Taskforce 2014).

But this attention from a growing set of foundations and investors is not the only driver of the attention to outcomes and impacts. Over the past two decades, nonprofits have come under increasing governmental scrutiny to demonstrate results and to improve accountability to the public, particularly when they are involved in public sector contracting (Krauskopf and Chen 2010; Smith 1999; Smith and Lipsky 1993). The Government Performance and Results Act of 1993 led many state agencies to develop milestones and adopt performance-based contracts with service providers, including nonprofit agencies, as part of broader efforts to “reinvent government” (Behn 2001; Poister 2003, xv). In a report to the U.S. Congress, and motivated by concerns raised by the U.S. Senate Finance Committee in the wake of the Sarbanes-Oxley legislation, the independent Panel on the Nonprofit Sector recommended that every charitable organization should “provide more detailed information about its operations, including methods it uses to evaluate the outcomes of its programs, to the public through its annual report, website, and other means” (Panel on the Nonprofit Sector 2005, 5, 37).

The most recent manifestation of this movement towards performance-based public funding is known as pay-for-success (PFS), an umbrella term for a wide range of mechanisms that typically involve partnerships between governments and private investors to finance social service providers based on performance. Among the mechanisms that have attracted the most attention are social impact bonds (SIBs) and development impact bonds (DIBs), in which private investors provide upfront capital for a social program, nonprofit organizations deliver it, and the government repays the investors with a rate of return only if the program meets prespecified performance targets.⁴ Early research suggests that SIBs are best suited to a fairly limited set of circumstances—where there is clear evidence of performance, coupled with a potential for substantial savings to public agencies—with much

of the upfront capital coming from philanthropy rather than private investors (Berlin 2016; Chamberlin 2018; Rangan and Chase 2015). Better evidence on the effectiveness and suitability of these mechanisms will become available over the next decade; as of early 2018, over one hundred impact bonds had been launched worldwide in sectors as diverse as health care, employment, agriculture, education, and social welfare, with results expected in over the next several years (Dear et al. 2016; Gustafsson-Wright et al. 2017; Shumway, Segal, and Etzel 2018). What is clear, however, is that the future of PFS will hinge on robust performance measurement—on setting metrics that accurately capture desired social outputs or outcomes, and in negotiating performance benchmarks among the contracting parties.

Performance measurement has a considerably longer history in the field of international development, where it is generally divided into impact evaluation (IE), and monitoring and evaluation (M&E). The former, as the nomenclature suggests, typically refers to the assessment of end results or impacts, while the latter is oriented towards informing midcourse correction by tracking the progress of a project or program. Since the 1950s, the international development field has seen various approaches to IE and M&E come in and out of fashion, with logic models and frameworks having become ubiquitous in the past forty years (Roche 1999, 18–20), diffused by a global industry of international development professionals, particularly consultants and evaluators employed by bilateral aid agencies and multilateral development institutions.

The use of logic models in particular, and program and project evaluation more broadly, has been spread in part by national governments and their development arms. These actors include bilateral aid agencies such as the United States Agency for International Development (USAID) or the United Kingdom's Department for International Development (DfID) and their counterparts elsewhere, as well as by multilateral institutions such as the World Bank and various agencies of United Nations. A more elaborate variant of the logic model, known as "logical framework analysis" (LFA), grew in popu-

larity among bilateral government agencies through the 1980s and 1990s, having been originally developed for USAID in 1969 (e.g., AusAID 2005; Commission of the European Communities 1993), more than a decade before its widespread adoption by nonprofits and their funders in the United States. The performance discourse in international development is often cast in terms of “results-based management” or “development effectiveness,” with a focus on setting clear goals and targets, identifying a causal logic model or results chain, and measuring results as a basis for continuous improvement (Morgan and Qualman 1996; Roduner, Schläppi, and Egli 2008).

More recently, bilateral aid agencies have sought to find ways of aggregating or assessing the sum total of results from their support to thousands of projects, programs, and civil society organizations worldwide. Their interests in doing so are twofold: to respond to accountability demands from their own finance ministries and taxpayers who want to know the “value for money” of their aid, and to increase development effectiveness towards global targets such as the Millennium Development Goals and the now more ambitious and ambiguous Sustainable Development Goals. These motivations were originally codified in the Paris Declaration on Aid Effectiveness, an international agreement signed in 2005, and the Accra Agenda for Action in 2008, both of which emphasized the need not only to measure results but to do so in a way that reflects national priorities and helps build country systems for managing development.⁵

The spread of logic models and impact measurement cannot, however, be attributed to foundations and government agencies alone. In the mid-1990s in the United States, for example, a number of national-level nonprofit organizations, particularly industry support groups, began to advocate for outcome and impact measurement while also developing tools for nonprofit managers. Among the most visible proponents was 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, supported by a series of resources designed to assist agencies in outcome measurement. The United Way’s

rationale for doing so was twofold: “To see if programs really make a difference in the lives of people” and “To help programs improve services” (United Way of America 1996). Over the following decade, the Urban Institute developed a common outcome framework and a series of indicators for over a dozen categories of nonprofits, later launching a web portal in collaboration with a software provider to house survey instruments and other tools for performance management (Urban Institute 2006, 2016).

The evidence on whether outcome measurement has led to improved performance is mixed. A study of thirty leading US 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 (Sawhill and Williamson 2001). The United Way of America showed gains in effectiveness among 391 agencies that it surveyed in 2000. It reported that over three-quarters of its partner organizations found outcome measurement useful for communicating results and identifying effective practices, as well as for helping to improve service delivery of programs. But the results were not all positive, with about half of the responding agencies reporting that implementing outcome measurement had led to a focus on measurable outcomes at the expense of other important results and an overloading of the organization’s record-keeping capacity, and that there remained uncertainty about how to make program changes based on identified strengths and weaknesses (United Way of America 2000).

Funders, too, have a mixed record of using impact assessments and evaluation in their decision making. In an analysis of evaluation methods used by philanthropic foundations over a period of three decades, the late nonprofit historian Peter Dobkin Hall argued that such evaluations lacked rigor and that key decision makers were often indifferent to the findings (Hall 2005, 33). Others have suggested

that philanthropic giving is often motivated by “expressive” interests of donors and not necessarily by evidence of what works and what doesn’t (Frumkin 2006; Knutsen and Brower 2010; Mason 1996). Many foundations continue to struggle with how to integrate a range of measurement approaches into their decision making. And some skeptical practitioners have suggested that while outcome measurement appears to be “a good tool to help funders see what bang they’re getting for their buck” (Glasrud 2001: 35), it runs the risk of being counterproductive when it draws precious resources away from services and reflects more of an obsession with expectations of funders than an interest in actually improving services and results (Benjamin 2008; Edwards and Hulme 1996a).

The practice of measurement is similarly mixed in the arena of impact investing. A survey of 146 impact investors conducted by J. P. Morgan and the Global Impact Investing Network found that while most investors lamented the lack of quality data, capacity, and methods for assessing the social performance of their investments, they were nonetheless satisfied with the social performance of their portfolios, as measured primarily by outputs rather than outcomes (J. P. Morgan and GIIN 2015). A more detailed follow-up survey found numerous improvements: 77 percent of 169 respondents claimed to be measuring outcomes rather than just outputs (although what outcomes were actually measured was not reported), and about one-quarter reported that the industry had made significant progress in improving investor understanding of impact measurement practice and in greater sophistication of tools and frameworks. Investors also identified a number of key challenges facing the industry, with about half expressing concern with a fragmentation of approaches to impact measurement and management, and one-third noting significant challenges to integrating social and financial management decisions, as well as on transparency about impact performance (Mudaliar et al. 2017).

These experiences of practitioner communities suggest that while social performance measurement remains at an early stage of

development, it is developing rapidly and perhaps even beginning to converge on common practices. Three broad convergent trends are particularly notable over the past decade. The first is a spurt of innovation in methods and tools for making measurement more useful and timely for decision making. This shift is captured by one of the pioneers in the impact investing field:

Metrics and evaluation are to development programs as autopsies are to health care: too late to help, intrusive, and often inconclusive. . . . [W]e set out to build a performance management process that would allow us to refine our thinking, change course, and diagnose problems before they became too significant. (Trelstad 2008, 107)

In the language of international development, this trend emphasizes the “monitoring” in “monitoring and evaluation” efforts. Nearly every large international development NGO has sought to reshape its M&E department to focus less on postprogram evaluation and more on mid-course correction and learning. And, there has been a flurry of efforts among impact investors and development evaluators to create “client-centric,” “lean,” and “right-fit” measurement approaches suitable for early stage and midstage enterprises with limited resources and capacity (Acumen 2015; Dichter, Adams, and Ebrahim 2016; IPA 2016; McCreless 2015). Ideally, of course, organizations need both—measures that help them refine and improve what they do on a day-to-day basis, and measures that help them keep sight of their long-term goals.

The second and closely related trend has been among providers of capital. One survey of 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” (Kramer et al. 2007, 5). Since then, many large funders have built up strong evaluation departments and measurement methodologies as part of a deeper commitment to strategic and outcome-oriented philanthropy (Brest 2012; Twersky, Buchanan, and Threlfall 2013; Weinstein 2009).

Over the past decade, the emerging impact investing industry has begun to build a measurement infrastructure characterized by standardized metrics, mapping and categorization initiatives, and third-party analytics and ratings systems (Mudaliar et al. 2017; Social Impact Investment Taskforce 2014). And the public sector, which has long used performance-based contracting, is showing renewed interest in its latest variation, pay-for-success contracts.

The third and deeper trend is a growing rationalization and marketization of the sector as a whole (Eikenberry and Kluver 2004; Hwang and Powell 2009; Mair and Hehenberger 2014; Powell, Gammal, and Simard 2005; Smith and Lipsky 1993). This rationalization is apparent in the hiring of professional managers and the widespread adoption of formalized practices typically associated with business such as strategic planning and independent auditing. Nonprofits and social enterprises are increasingly recruiting graduates of MBA programs while also seeking the services of strategy consulting firms. As part of this rationalization of the sector, they are also increasingly seeking to measure and even quantify their performance (Brest 2012; Bromley and Meyer 2014; Ebrahim 2003b; Frumkin 2002; Tuckman and Chang 2006; Young and Salamon 2002).

These three broad trends—a spurt in innovation around real-time feedback and midcourse correction, growing capital markets for social performance, and a steady rationalization of the sector—all point to the increasing institutionalization of performance measurement in the social sector. They also portend a growing need for managers to be more deliberate about what they seek to achieve and how to measure progress towards it. These are challenges of both measurement and strategy.

Linking Measurement to Strategy

I turn now to the core practical challenges that emerge from the above discussion. How can leaders of social sector organizations be more deliberate about performance measurement and management? This

is a question of strategy. I use the term *strategy* to mean *how an organization seeks to achieve long-term performance*.⁶ The role of performance measurement and management is to *support the design and execution of strategy, as well as its improvement over time*. The key strategic challenges of performance can thus be captured in terms of three foundational questions facing every organization:

1. What does my organization seek to achieve?
2. How will we bring about that change?
3. How will we hold our feet to the fire?

The first question captures the organization's value proposition, the second its operational or social change model, and the third its own terms of accountability. Figure 1.2 depicts these components of social sector performance, which I discuss briefly below. The concepts and terms discussed here will be familiar to many; my intent is to get all readers on the same page as quickly as possible before proceeding to a more complex and nuanced examination of performance measurement. For those less familiar with the concepts described below, I provide further resources in the appendix to this book.

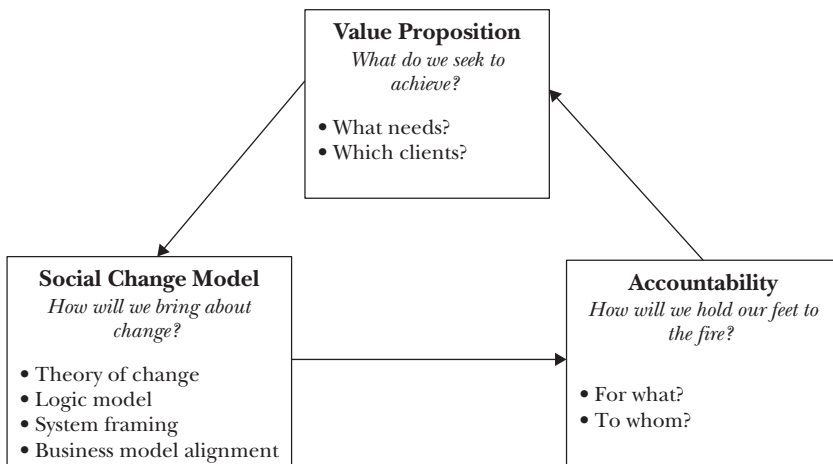


FIGURE 1.2 Core Components of Social Performance

Value Proposition

At the heart of any strategy lies a value proposition (Moore 2000; Porter 1980; Rangan 2004). It answers a fundamental question: *What do we seek to achieve?* It is outward looking, specifying the value to society that the organization seeks to create in terms of two questions:

- What problem or needs will be addressed?
- Which clients or population will benefit?

A good social value proposition provides clear answers to the above questions and is thus ultimately measurable. While most social sector organizations have mission statements, many are too broad to qualify as a value proposition. Consider an organization serving the homeless. In this book, we will examine the case of Miriam's Kitchen, which changed its mission from a vague statement—"to provide individualized services that address the causes and consequences of homelessness in an atmosphere of dignity and respect both directly and through facilitating connections in the Washington, DC community"—to a more concise and measurable value proposition—"to end chronic homelessness in Washington, DC." This much simpler statement clearly identifies both the social problem and a well-defined target population. The organization targeted an especially difficult subset of the homeless population: the chronically homeless, defined as individuals who have been continuously homeless for at least one year and have experienced four or more episodes of homelessness in the prior three years. The clarity of this value proposition provides a starting point for designing an operational model to address the problem.

Social Change Model

While a value proposition looks outward to the value created for a segment of society, a social change model looks inward to the organization's operations, posing the question: *How will we bring about change?* Any model of social change is made up of multiple parts:

- theory of change

- logic model or results chain
- system framing
- business model alignment

A *theory of change* (or causal logic) explicitly articulates how an organization's interventions will address the social problem specified in its value proposition. Put another way, it specifies the cause-effect relationships or pathway through which actions are expected to generate results (Center for Theory of Change 2014; Keystone 2009; Weiss 1995). A good theory of change can be stated as a series of hypotheses or “if-then” statements—*If* we do a, b, and c, *then* we expect results x and y to occur. For example, an organization serving the homeless may engage in many activities including providing meals, overnight shelter, substance abuse counseling, health services, or even lobbying city hall to allocate resources for permanent housing. Its theory of change must articulate the pathway through which these activities can be expected to reduce homelessness in the city. The cases in the next four chapters of this book illustrate a variety of theories of change, ranging from fairly straightforward pathways supported by substantial scientific evidence to complex ones for which research remains sparse. A challenge facing many social sector organizations is to surface their unquestioned assumptions about cause and effect in order to develop an explicit and testable theory of change. The appendix to this book provides several resources to help organizations develop their theories of change.

A theory of change is further operationalized with the help of another tool—a *logic model* or *results chain*—that specifies the key inputs and activities required to generate the desired outputs and outcomes (see Figure 1.1 above). The primary value of a logic model lies in its role as a disciplining mechanism, forcing managers and staff to articulate the steps in their interventions and the measurable results they expect to see along the way. Each arrow in a logic model can thus be understood as a hypothesis amenable to testing: Do the activities produce the expected outputs? Do the outputs lead to anticipated

outcomes? A well-developed logic model can be useful not only for identifying performance metrics along each step in the chain, but also for uncovering opportunities for improving performance.⁷ As we shall see throughout this book, however, developing a useful logic model is often complicated by the uncertainties about cause and effect inherent in many social change efforts.

The third component of a social change model is a *system framing* that positions the organization's interventions within a broader social or institutional context. Take, for instance, an ambulance service that responds to emergencies and delivers patients to a hospital. Even though such a service may operate as an independent enterprise, its effectiveness ultimately hinges on plugging into a larger health-care system so that it can access hospitals and deliver patients to facilities that can meet their needs. Similarly, an organization serving the homeless with meals or overnight shelter is part of a larger ecosystem of actors attempting to address urban poverty. The cases in chapters 4 and 5 of this book illustrate the usefulness of a system framing, especially for organizations that seek to influence public policy or to collaborate with other actors in order to achieve their social goals.

Finally, an effective social change model requires *alignment with its business model* in order to minimize conflict between social and financial performance goals (Ebrahim, Battilana, and Mair 2014). In a highly competitive funding environment, it is not uncommon for organizations to accommodate the wishes of their funders even if that means compromising their social objectives. Even social enterprises that generate revenue directly from their customers can lose sight of their social goals in their efforts to generate revenue, a risk referred to as *mission drift* (Fowler 2000; Jones 2007; Weisbrod 2004). For instance, a health-care business serving low-income communities will constantly be tempted to also serve middle-income populations in order to bolster its bottom line. On one hand, doing so may enable it to expand its services to more low-income communities. On the other hand, it may succumb to mission drift by reducing its services to the poor in order to become more financially successful. This dilemma is

further exacerbated by the fact that financial performance is easier to measure and audit than social performance. As such, a key strategic function of performance measurement in the social sector is to ensure that financial performance goals support, rather than overtake, social performance goals.⁸

Accountability

The final pillar of social sector performance is accountability—how the organization plans to hold its own feet to the fire. I focus here on the internal terms of accountability that the organization sets for itself, rather than on the expectations of external actors such as funders. There are two main dimensions of accountability that an organization's leaders must clarify as part of their strategy (Behn 2001; Ebrahim 2016; Kearns 1996; Mulgan 2000; Najam 1996a):

- accountability *for what*?
- accountability *to whom*?

Specifying *for what* the organization is accountable requires identifying the social results (outputs and outcomes) that are most important to its performance. Unlike in the business world where there is general agreement about standard measures of success, accountability in the social sector is complicated by a lack of common standards or benchmarks for measurement, and the general difficulty in comparing social performance across organizations. The task thus falls on leaders and managers to develop metrics that not only capture the organization's value proposition but also can realistically be delivered by the organization's model of social change. For example, it does not make much sense for an emergency medical service to measure long-term health outcomes when all it can reasonably deliver is a short-term response such as delivering a patient to a hospital. On the other hand, without ambitious outcome targets such as ending chronic homelessness, an organization that serves the homeless may fail to achieve its full potential.

Moreover, it is important for organizations to develop internal clarity about *to whom* they owe accountability. Even when an organization

has clear measures of its performance, these measures may not satisfy all of its key stakeholders, particularly if their interests are divergent or in conflict.⁹ An important task for leadership is to prioritize among various stakeholders and then develop mechanisms through which those central stakeholders can hold them to account.¹⁰ In the absence of such mechanisms, organizations run the risk of defaulting to the interests of their most powerful stakeholders rather than those who are most vulnerable and most important to their missions. Typically, the most powerful stakeholders tend to be donors or investors rather than beneficiaries or clients, because funders can threaten to withhold funding whereas clients often do not have such an exit option or sanctioning mechanism (Hirschman 1970; Najam 1996a). This condition, prevalent in the social sector, is often described as strong “upward” accountability but weak “downward” accountability (Ebrahim 2003a; Edwards and Hulme 1996a), or even as “multiple accountabilities disorder” (Koppell 2005). The case studies in this book show that there are many avenues for strengthening downward accountability, for example, by systematically collecting client feedback on goods or services, allocating decision-making rights to clients in the work of the organization, and building the capacity of clients to engage in policy advocacy. We shall explore these mechanisms through the case studies that follow.

In sum, a key task for leaders is to specify how they will hold their organization’s feet to the fire—by clarifying *for what* they are accountable and prioritizing *to whom* they owe accountability. This is the “intelligent forms of accountability” referenced in the opening to this chapter. By setting their own terms of accountability, social sector leaders can establish performance goals for themselves rather than merely being on the receiving end of the accountability demands of funders or other stakeholders. This is an accountability based on strategy rather than compliance.

These three components of social sector performance—value proposition, social change model, and accountability—constitute a virtuous cycle within an organization. A value proposition clarifies the organization’s social purpose, its social change model puts that

purpose into action, and accountability closes the loop by judging whether that purpose is being achieved.

A Contingency Framework for Measuring Social Performance

With the brief introduction to social sector strategy above, I turn now to the questions of performance measurement at the heart of this book: *What results should social sector organizations measure? What kinds of performance measurement and management systems should they build?*

Conventional wisdom in the social sector suggests that one should measure results as far down the results chain as possible—measuring not only the activities and outputs produced by an organization, but also its long-term social outcomes. This expectation is based on a normative view that organizations working on social problems should be able to demonstrate results in solving societal problems. Yet it is worth pausing to consider whether, and to what degree, such measurement makes sense for all types of social sector organizations. A crude typology of the social sector in the field of poverty alleviation distinguishes among at least three broad types of activities (Najam 1996b; Vakil 1997):

- Emergency relief—activities that address urgent survival needs, such as food and temporary shelter, as well as disaster, crisis and conflict mitigation
- Service delivery—activities that address basic needs, such as education, health care, longer-term shelter, community development, employment and income generation
- Policy and rights advocacy—activities that address structural issues related to rights, public policy, regulation, and societal norms

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 complicated activity requiring highly sophisticated coordination and logistics management capabilities, it is focused on meeting immediate survival needs rather than long-term development outcomes. 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? Take the example of an immunization campaign against a contagious disease. A key output measure of an immunization campaign is the number of people vaccinated, against polio for instance. Although each person immunized is important, the most significant impacts are only achieved once a certain “herd immunity threshold” is reached, about 80–86 percent of a population in the case of polio, as this makes it possible for the disease to be eradicated. The metrics are typically expressed as outputs (number or percentage of people vaccinated) and outcomes (declines in illness) in order to get at impacts at scale (prevention, containment, or eradication of a disease). While the causal logic in preventing or eradicating a disease may be relatively well established, executing it turns out to be extremely complicated—it requires having not only a proven vaccine and the technology for distribution but also strategies for community organizing, for addressing cultural norms and fears that may limit immunization uptake, for coordinating public health workers, and for replicating these efforts at scale. Only two diseases, smallpox and rinderpest, have ever been eradicated through immunization campaigns.¹¹

Outcome measurement turns out to be uncommon in the social sector for the simple reason that organizations have the most control over their immediate activities and outputs, whereas outcomes are often moderated by events beyond their organizational boundaries. An 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 nonprofits. Connecting individual-level outcomes to societal-level outcomes, 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.

The link between outputs to outcomes turns out to be even more complex in cases of policy advocacy and rights-based work. Well-known organizations like Amnesty International, CARE, Human Rights Watch, Oxfam, and others aim to address the root causes poverty and injustice through their policy-based work. But their abilities to measure long-term results, and to attribute those results to their interventions, are severely limited. Their advocacy campaigns may achieve many successes—for example, in getting the attention of public officials to the rights of children, women, political prisoners, and others; in forming national-level constituencies and coalitions; and in improving public awareness about marginalized groups. However, these measures are largely about activities, processes, and outputs, rather than about ultimate outcomes on rights. This focus on activities and process is due, at least in part, to the complex nature of causality when it comes to shaping public policy and social norms. It is also about the difficulty in attributing any one organizational intervention, among an ecosystem of interventions, to societal outcomes.

These preliminary observations enable us to relate performance measurement to the type of social sector activity:

- Performance in emergency and relief work can be measured in terms of inputs, activities, and outputs.

- Performance in service delivery work can be measured in terms of activities, outputs, and only sometimes outcomes.
- Performance in advocacy and rights-based work can be measured in terms of outputs and possibly “influence” (an interim outcome).

Perhaps what is most surprising about these initial observations is that only one category of intervention can readily be measured at the outcome end of the logic chain—service delivery activities for which the causal link between outputs and outcomes is well established and where the organization can exercise sufficient control over its environment to ensure fidelity to the causal model. In contrast, rights-based and advocacy interventions, despite aiming to address root causes of social problems, cannot easily demonstrate a causal link and are thus typically left with measuring outputs. These constraints on measuring performance also present a dilemma for accountability: many funders increasingly want to see measures of outcomes, whereas it may not be feasible for many organizations to measure performance beyond outputs.

In short, outputs don’t necessarily translate to outcomes, and individual outcomes don’t necessarily translate to societal outcomes. 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. But when should it step forward into the domain of outcomes and impacts?

Two Contingencies: Causal Uncertainty and Control Over Outcomes

Addressing this question requires a deeper understanding of the causal models of social change. In an important conceptual advance, Rogers (2008) differentiates among three categories of causal logics: simple, complicated, and complex (a distinction she credits to Glouberman and Zimmerman [2002] and which is also closely related to Snowden [Snowden and Boone 2007, Snowden and Kurtz 2003]). Most organizations employ a *simple* logic model, which assumes a

linear cause-effect relationship among variables as a basis for setting performance targets and for evaluating success. She suggests that such models are best suited to interventions that are tightly controlled, well understood, and homogenous. Arguably, such interventions presuppose rational behavior and are amenable to technocratic control. *Complicated* logic models are more useful for interventions that involve multiple actors in implementation, multiple causal strands that must operate simultaneously in order to achieve success, or alternative causal strands where there are different pathways to the same result.

Complex logic models are distinguished by their nonlinearity, where the causal logics and outcomes are insufficiently understood, interactive, and multidirectional. They may involve disproportional impacts where small effects can magnify through reinforcing loops or at critical tipping points. Rogers further suggests that complex causal models are developed through processes of recursivity, requiring feedback loops for convergence and negotiation among multiple actors, or through an emergent process where patterns are gradually identified through ongoing interaction and analysis. She cautions that the high degree of uncertainty associated with complex causal logics “can lead managers and evaluators to seek the reassurance of a simple logic model, even when this is not appropriate” (Rogers 2008, 45).

Framed another way, simple logic models seem best suited to conditions of rationality within closed systems where knowledge of cause and effect is well established and where organizations exercise considerable control over each step in the results chain. Complicated logic models, however, may be more useful for assessing interventions that play out in rational but open systems where environmental influences are more difficult to contain, requiring special efforts to integrate various interventions. And complex logic models may be appropriate for organizations operating in open systems, but not necessarily fully rational or technical ones, that require greater recursivity and coordination among external actors for identifying and assessing results and their causal pathways.¹²

However, this spectrum of logic models—from simple to complicated to complex—merges two distinct factors that are useful to ex-

amine separately: *uncertainty* in cause-effect knowledge for addressing a social problem and *control* over the activities and conditions necessary for producing long-term outcomes. For social interventions, it is necessary to consider these two contingencies in order to identify the types of results an organization can reasonably be expected to produce.¹³

The first of these variables, uncertainty about cause and effect, has been a long-running theme in the management and accounting literature, especially as shaped by contingency theory (Chapman 1997; Chenhall 2003, 2006; Galbraith 1973; Hopwood 1980; Thompson 1967; Thompson and Tuden 1959). One key insight of this research is that high uncertainty in cause-effect relations makes it difficult to specify which behaviors are necessary for achieving desired outcomes prior to task completion, and even to evaluate the appropriateness of behaviors after task completion. This insight substantiates the reservations of evaluation researchers about the usefulness of the *ex ante* specification of cause-effect models under high causal uncertainty. Under these conditions, there may exist a temptation to use overly simplified linear logic models for program evaluation, which can be valuable for gaining organizational legitimacy but may also create an unintended focus on narrow activities or rationales for an intervention (Blalock 1999, 136–37; Lindgren 2001).

The second contingent factor focuses on the degree to which an organization controls all of the activities and conditions necessary for delivering long-term outcomes. Exercising control over social outcomes is complicated by the fact that it typically requires sequencing and coordination of multiple interventions, or even multiple actors, in an ecosystem. In organization theory, such “interdependence” is seen as a necessary consequence of the division of labor in joint activities (Victor and Blackburn 1987). In their seminal work on power and control in organizations, Pfeffer and Salancik (1978, 40) define interdependence as a phenomenon that “exists whenever one actor does not entirely control all of the conditions necessary for the achievement of an action or for obtaining the outcome desired from the action.”¹⁴ The strategic challenge for social sector organizations thus lies in how to integrate or coordinate the constellation of activities and actors

that produce collective outcomes. In this book, I explore a number of mechanisms of coordination for increasing control over outcomes.

These two contingent factors form the axes of the contingency framework depicted in Figure 1.3. First consider the vertical dimension of the matrix, which refers to the causal uncertainty underlying an intervention. When there is low uncertainty, the relationships between cause and effect are relatively well understood and stable, and one has fairly complete knowledge of how a set of actions will lead to outcomes. In the social sector, such clarity is most likely to exist for interventions involving a linear or “simple” logic model (Rogers 2008). In contrast, high uncertainty refers to interventions where knowledge about causal pathways is relatively incomplete, and where cause-effect relationships are believed to be “complex” in that they may be nonlinear, interactive, and emergent.

In more commonly used practitioner language, the vertical dimension of Figure 1.3 refers to an organization’s theory of change, which

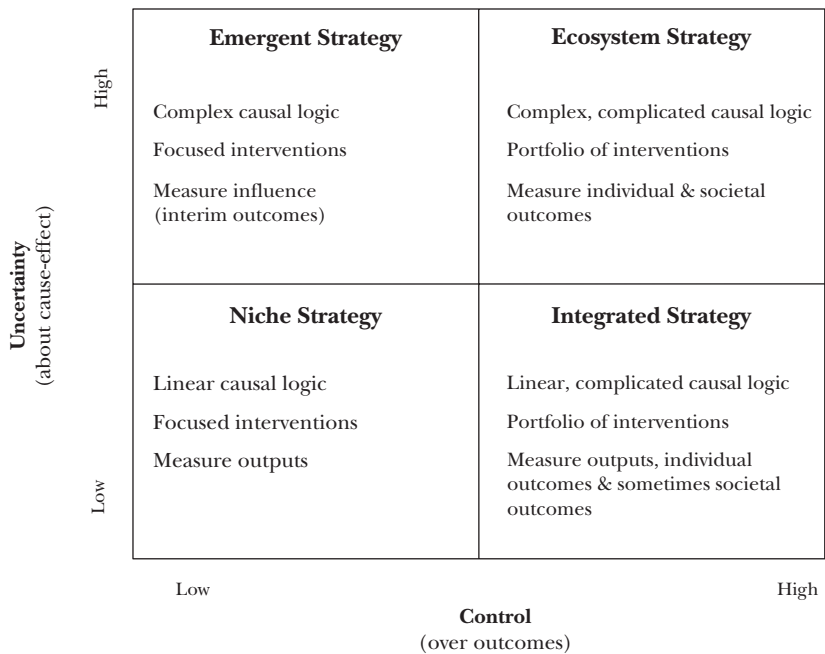


FIGURE 1.3 A Contingency Framework for Social Performance

articulates the causal logic or pathway through which a set of interventions is expected to lead to a long-term goal. How complex is the relationship between cause and effect, and how well is it understood? The bottom half of the matrix refers to settings where the theory of change is well established and is typically linear or simple. The upper half of the matrix applies to settings where the theory of change is complex, such that cause-effect relationships are only weakly understood, and where there may be multiple causal factors at play.

We can apply this reasoning to the examples above of organizations engaged in the delivery of emergency and basic services after a natural disaster such as an earthquake or hurricane. They operate on a clear and well-established causal logic: get shelter, food and water, and medical care to people facing a crisis in order to avert further disaster. Or provide access to clean water and sanitation to prevent outbreaks of waterborne illness and improve human health. In order to deliver aid quickly, they must acquire and stock critical supplies such as blankets, tents, food, water, and medical supplies, while also transporting and distributing those supplies and ensuring that they are delivered equitably and safely on the ground. The supply chain management capabilities needed to deliver such aid can be enormously complicated, but the basic intervention logic is fairly well understood. Similarly, immunization campaigns follow a well-established causal logic based on evidence about the vaccination rates required in order to reduce or eliminate a disease, even though the operational strategy can be very complicated. Such interventions fall in the bottom half of the matrix, as the knowledge about causal mechanisms is fairly well understood. Under such conditions, output measures are useful for assessing performance.

On the other hand, consider efforts by organizations to influence public policy or foster democratization and human rights, such as by Amnesty International, Oxfam, or various civil liberties groups. Their abilities to attribute any long-term changes in public policy or social norms to their interventions are severely limited. These are settings with “complex, foggy chains of causality” where “events evolve rapidly and in a nonlinear fashion, so an effort that doesn’t seem to

be working might suddenly bear fruit, or one that seemed to be on track can suddenly lose momentum” (Teles and Schmitt 2011). Under these conditions, it would be foolish to attribute outcomes to any one organizational intervention. While an organization may influence or *contribute* to a policy or attitudinal change, it is more difficult to *attribute* that change primarily to its interventions. In such contexts, it is hard to know how much of a difference a particular intervention makes. Causality would be hard to establish, although one could make credible claims about “influence” on a set of outcomes (an interim outcome).

To be sure, attribution and contribution are both about establishing a causal relationship between interventions and an outcome. But they differ in the degree of uncertainty about cause-effect knowledge, and the methods for testing causal claims. A claim of attribution requires specification of: (1) a well-defined intervention or “treatment” applied to a unit or treatment group; (2) an observable outcome; (3) a counterfactual value, that is, the outcome expected in the absence of the intervention; and, (4) a means of assigning units or groups that receive the intervention and those that do not (treatment and control groups).¹⁵ Satisfying these four conditions is important for testing claims of causal attribution; doing so is feasible for many social interventions involving the delivery of goods and services such as medicines, agricultural inputs, or even education and training. The evaluation methods ideally deployed in such contexts are experimental, such as randomized control trials, which enable the assignment of some kind of control group for comparison (Banerjee and Duflo 2009; Imbens and Rubin 2015; White 2006).

But there are many interventions and contexts where these conditions cannot be satisfied. For instance, consider an organization aiming to influence national health-care or labor policy. It may have a well-defined intervention (a lobbying strategy with specific arguments and tactics that target specific legislators) and an outcome it seeks to achieve (passage or changes to proposed legislation). Similarly, an organization seeking to influence negotiations in a civil conflict may have a strategy for influencing key players and a clear outcome in

mind. But in neither of these settings is it typically feasible to establish treatment and control groups. These difficulties do not obviate the need for making and testing causal claims. But they require a different toolkit better suited to assessing causality in complex nonlinear contexts, such as methods of “process tracing” used by political scientists and “contribution analysis” developed by evaluators (Bennett and Checkel 2015; George and Bennett 2005; Kane et al. 2017; Lemire, Nielsen, and Dybdal 2012; Mayne 2011, 2012). In Chapter 7, I offer a series of resources for carrying out evaluations using attribution-based and contribution-based methods, and I revisit the conditions to which each is best suited.

Moreover, the distinction between attribution and contribution is not merely methodological or technical. It also has implications for strategy. Any organization that chooses to work collaboratively with other actors must confront the problem of how to “take credit” for its role when communicating with funders. Measurement focused primarily on attribution runs the risk of undermining the collective efforts of actors in an ecosystem when it incentivizes them to seek credit and funding for individual behavior (a zero-sum game) rather than to produce interdependent results (a mutual-gains game). In such contexts, managers and funders alike are better off identifying a constellation of factors that jointly affect a social problem (contribution) rather than obsessing about how to isolate the causal role of each factor or assigning weights to those factors based on statistical correlation (attribution). This distinction is important because it forces managers to be honest about the limitations of their claims of cause-effect, while still taking action under uncertainty. I revisit this distinction in chapters 4 and 5 in order to examine the challenges of measuring results when working with other actors in an ecosystem.

Now consider the horizontal axis of the matrix in Figure 1.3, which refers to an organization’s control over outcomes. Under conditions of low control, an organization focuses on delivering a highly specific task or output that, on its own, is insufficient for addressing a social problem. Such an output may provide a necessary service—

such as meals or emergency shelter for the homeless, or job training for the unemployed—but it is unlikely to be sufficient for delivering an outcome such as ending homelessness or providing stable employment to a target population. In contrast, an organization may choose to increase its control over outcomes by combining multiple interventions—such as job training in combination with counseling, job placement, and postplacement coaching—that together are more likely to address the social problem. In practitioner terms, this is the “operational strategy” of the organization: the organization’s leaders must choose whether to adopt a highly focused strategy that delivers a limited set of outputs with high quality, or to adopt a portfolio strategy that brings together a wide range of activities that, in combination, are likely to increase its control over outcomes. This is a strategic organizational choice.

Taking into account these contingencies, Figure 1.3 identifies four broad or “ideal types” of strategies that can be undertaken by organizations: niche, integrated, emergent, and ecosystem strategies. These four types should not be read as a normative continuum, but instead as a reflection of the deliberate strategic choices made by social sector actors.

A *niche* strategy is an appropriate choice when there is good knowledge about cause and effect but where control over outcomes is limited. There are two basic kinds of value propositions that fall into this quadrant. The first involves interventions where the desired results are outputs rather than outcomes. For example, emergency and relief services that provide temporary shelter or food aim to deliver a very specific set of outputs. These outputs are valuable in themselves, despite the fact that they do not aim to affect longer-term outcomes for individuals. They may be assessed in terms of the quality and efficiency (outputs/inputs) of the goods or services they provide.

A second kind of value proposition favoring a niche strategy is when evidence about cause and effect is sufficiently well established such that output measures can serve as a reasonable proxy for outcomes, even in the face of low control over those outcomes. For ex-

ample, a medical intervention that treats children for intestinal worms has been shown to decrease school absenteeism in the developing world and thus increase employment earnings of those participants over time (Baird et al. 2011; Miguel and Kremer 2004). These findings were based on a randomized evaluation of a government deworming program conducted in seventy-five primary schools in a rural district of Kenya, involving over thirty thousand children. A representative sample of about 7,500 respondents was assessed ten years after the intervention. Where such research can establish confidence in the causal links between the outputs and desired long-term outcomes, it is sufficient for the implementing organizations to execute a niche strategy and focus on measuring outputs (numbers of children dewormed) and the quality and efficiency with which those outputs are delivered. Admittedly, such instances remain rare in the social sector, as they typically require meeting the conditions for a randomized control trial.¹⁶

An *integrated* strategy is characterized by good knowledge about cause and effect, but a need to coordinate among multiple activities or actors in order to increase control over outcomes. Take the example of a job training program that provides technology skills to unemployed youth. The organization may deliver a high-quality training program necessary for helping its clients climb out of poverty. But it does little to influence the actual employment opportunities for those youth or to address the problem that many youth are likely to lose their jobs within several months of being placed in them. To have real impact, additional interventions are needed, such as creating a job placement pipeline with employers, coaching the youth when they have difficulty retaining jobs, and working with employers to provide support when youth underperform (e.g., Chertavian 2012). The niche intervention of a job training program is insufficient and must be delivered in coordination with other upstream and downstream interventions. These additional interventions may be undertaken by the same organization as part of a vertically integrated strategy, or they may be done in collaboration with other organizations. Either way, coordination of the interventions is essential to delivering outcomes.

An *emergent* strategy, in contrast, is appropriate under conditions of relatively poor knowledge about cause and effect and little control over outcomes. A policy advocacy organization, for instance, faces a dual challenge in its efforts to influence public policy: a need to quickly adapt its strategy in order to take advantage of new policy windows or a changing political landscape, and a lack of control over the policy-making process and policy makers. Its strategy is necessarily emergent in that it must learn to respond quickly to changing conditions (Mintzberg 1978), all the while staying focused on its advocacy goals. Performance measurement is especially messy in such settings due to the nonlinear and interactive nature of causality. The organization might claim to have influenced a public policy but could only rarely provide compelling evidence that it was actually responsible for a policy change. Another example lies in the field of higher education. While there is well-established evidence that individuals with college degrees are able to get better-paying jobs than those without such education, the missions of many colleges and universities to advance knowledge, to build an engaged citizenry, and to contribute to society are more difficult to measure due to the cause-effect uncertainties involved, particularly in the liberal arts (Faust 2013; Rosenman 2013). Here, too, modest claims of influence on such outcomes are plausible.

Finally, an *ecosystem* strategy must grapple with complex cause-effect pathways that are poorly understood, while simultaneously seeking to increase its control over outcomes. It does so by orchestrating the interventions of multiple actors in an ecosystem in order to produce outcomes that no single actor could produce alone (Bradach and Grindle 2014; Kania and Kramer 2011; Montgomery, Dacin, and Dacin 2012; Weber 2003; Wei-Skillern and Marciano 2008). Consider a homeless-service provider that feeds its clients a warm breakfast and is focused on immediate outputs (meals), but any longer-term outcomes (getting people out of homelessness) are typically beyond its control. Other organizations in the same city may be providing substance abuse counseling, job training, or shelter. But none can independently reduce or end homelessness. To address this problem, many

communities across the United States are increasingly coordinating services by first placing clients in housing and then matching services to their needs. The cause-effect relationship between each intervention and the outcome of reducing homelessness is largely unknown, but their collective impact is measurable. Success hinges on better coordination of previously isolated actors and services within an ecosystem (e.g., Culhane, Metraux, and Hadley 2002; National Alliance to End Homelessness 2006; Tsemberis and Eisenberg 2000).

Taken together, the two axes of the framework embed social performance in relation to an organization's knowledge about cause-effect and its control over outcomes. The vertical dimension of the framework acknowledges the constraints on rationality in organizations, particularly for addressing complex social problems where causal knowledge remains limited. The horizontal dimension differentiates between two kinds of operational choices, where an organization may choose to deliver highly focused interventions or to assemble a portfolio of interdependent interventions for generating outcomes.

Implications of the Contingency Framework

The four types of strategies and their performance measures outlined in Figure 1.3 are not intended as a normative continuum or as mutually exclusive domains, but instead as ideal types reflective of the contingencies faced by social sector actors. This typology leads to a number of implications that I summarize here and explore further in the chapters that follow.

First, and perhaps most important, *it is possible to be high performing in each of the four quadrants of the matrix*, where high performance is defined as “the ability to deliver—over a prolonged period of time—meaningful, measurable, and financially sustainable results for the people or causes the organization is in existence to serve” (Performance Imperative 2015). The yardsticks for what constitutes “meaningful” results must necessarily vary with the contingencies in the matrix. Unless social sector leaders have a grasp of both contingent factors—the state of knowledge about cause and effect for the problem they are trying

to address, and their degree of control over outcomes—it will be impossible for them to determine appropriate measures of performance.

Second, the measurement of long-term outcomes is feasible under a limited set of circumstances—in the integrated and ecosystem quadrants of the framework. In such settings, seeking to attribute outcomes to a single intervention does not make much sense. To increase their control over outcomes, organizations face two strategic choices. They can either better integrate a series of interventions in-house (such as a vertically integrated job training program that does job training, job placement, and job retention coaching), or they can better orchestrate a network of service providers. Only in a limited set of circumstances are they likely to have sufficient control over the entire delivery chain to reasonably measure and take credit for outcomes.

Third, outputs can be a valuable measure of social performance, particularly under conditions of low causal uncertainty. For some interventions, such as emergency relief services, the desired results are best captured in terms of outputs. For others, where cause-effect knowledge is well established, outputs can serve as a reasonable proxy for outcomes. This does not, of course, mean that outputs are always a good measure of performance, but only that managers must be clear in their rationale when they use outputs.

Fourth, many organizations will not fit neatly into the quadrants of the matrix, with some traversing many of them. Complex international development organizations like Oxfam or CARE, for instance, are engaged in a portfolio of activities ranging from emergency relief to service delivery to rights-based work. They face the double challenge of measuring performance in each of these areas separately while also integrating across them in order to gauge their possible synergistic effects at an ecosystem level.

And finally, this contingency framing presents a dilemma for accountability: donors and investors increasingly want to see measures of long-term societal outcomes, whereas it may not be appropriate for many operating organizations to measure performance beyond outputs and possibly individual-level outcomes. In other words, there is a gap between what funders and the general public may desire from

social sector actors and what they can reasonably be held accountable for. Managers who focus on measuring outputs or influence (as in the niche and emergent quadrants) risk being seen as failing to be accountable to some of their stakeholders, failing to convince funders and citizens that they are making a difference in society, thus risking both their support and their legitimacy. At the same time, managers who seek to take credit for societal outcomes (as in the integrated and ecosystem quadrants) risk overreaching by claiming social changes that may be beyond their actual control and for which they are not solely responsible, potentially undermining their own credibility (Campbell 2002).

In sum, the contingency framework depicted in Figure 1.3 reflects the open systems nature of social change rather than treating it as a closed rational system. Only in a limited set of circumstances are social sector organizations likely to have sufficient clarity about cause and effect, and control over the results chain, to reasonably take credit for long-term outcomes. As a result, we might expect the role and nature of performance measurement and management systems to vary with each quadrant of our matrix. Using a diverse range of cases, the next four chapters of this book examine what performance systems might look like in each quadrant of the figure.

Building Systems for Social Performance: An Overview of Four Case Studies

The following four chapters provide in-depth case studies on each type of social change strategy identified in the matrix in Figure 1.3—niche, integrated, emergent, and ecosystem. I summarize each of the four cases below. Readers most interested in a particular type of strategy and its performance measurement system can proceed directly to the relevant chapter. Chapter 6 provides a synthesis of the findings and develops diagnostic frameworks for managers.

Each case study has three objectives: (1) to move from conceptualizing “what to measure” to examining its operational details, (2) to develop general insights on the type of performance measurement and

management system best suited to supporting each type of strategy, and (3) to consider the tensions and difficulties common to developing a robust performance system for each type of strategy. The cases have been selected with these three goals in mind. But perhaps what makes the cases most interesting is the continuous struggle within each organization to adapt its performance measurement system to its evolving needs. There is no such thing as perfection in these dynamic contexts, only constant improvement.

Chapter 2 examines the challenges of measurement common among organizations that adopt a *niche* strategy, where the relationship between cause and effect is relatively well understood and the intervention is highly focused. The main argument in this chapter is that, under such conditions, managerial attention is best focused on implementing the intervention with high quality—which requires a focus on short-term metrics of activities and outputs rather than on long-term outcomes. I develop this argument through the case of Ziqitza Health Care Limited, a social enterprise that is one of the fastest growing private ambulance services in India. Ziqitza's experience offers two general insights. First, it illustrates the conditions under which managerial attention is best devoted to tracking short-term outputs. Second, it highlights the role of *standardization* and quality control as a defining feature of an output-focused performance measurement system. The chapter also draws briefly on other examples of niche interventions such as postdisaster emergency response.

Chapter 3 explores the performance systems necessary for supporting an *integrated* strategy, where the relationship between cause and effect is relatively well understood but where multiple interventions must be combined in order to produce outcomes that are greater than the sum of their parts. Here, the core tasks of a performance measurement system is *coordination*, which involves prioritizing and sequencing a portfolio of activities and measuring their synergistic effects. I examine the experience of the Aga Khan Rural Support Programme India, looking at a pipeline of integrated interventions in rural agricultural and natural resource management that aim to in-

crease incomes of smallholder farmers. This portfolio of interventions includes agricultural credit, access to high-quality inputs such as seeds and fertilizers, farmer training, soil and water conservation interventions, and access to markets. I track the organization's challenges in measuring the performance of these interventions over a twenty-year period. This chapter concludes with other examples of integrated strategies, including an even more complex endeavor to assist people living in extreme poverty by BRAC in Bangladesh, and a pipeline of educational interventions developed by the Harlem Children's Zone in New York City.

Chapter 4 shifts the focus to an *emergent* strategy, where the relationship between cause and effect is complex and control over outcomes is very limited. Such organizations are best served by an *adaptation* system that enables quick iteration and adjustment of strategy in response to unexpected opportunities. I draw on a complex case of a global network organization called WIEGO—Women in Informal Employment: Globalizing and Organizing—which advocates for the rights of workers in the informal economy. Such workers include, for instance, street vendors, waste pickers, and domestic workers who typically lack social protections such as working condition standards and minimum wages. I explore the efforts by WIEGO to influence standard setting at the International Labor Organization (ILO). It would be difficult for WIEGO to demonstrate that it has improved the working conditions of informal workers, but it can reasonably claim to have “influenced” the ILO's standards and guidelines on measuring the informal sector, which in turn shape how national governments measure and support their informal sectors. WIEGO's performance system is characterized by constant adaptive learning: quick actions to take advantage of policy windows and recalibration of strategy based on new information, but tempered and guided by clear long-term goals. Such a performance system is not unique to WIEGO but is relevant to a diverse range of organizations engaged in changing social policies and norms. In closing this chapter, I draw parallels to two very different types of interventions—a campaign to prevent teen

pregnancy in the United States and a leadership training program in Burundi in the aftermath of a violent ethnic civil conflict.

Chapter 5 turns to an *ecosystem* strategy, where not only are cause-effect relationships complex but also where multiple organizations collaborate in order to increase their control over outcomes. In such settings, causal attribution is nearly impossible because organizations cannot reliably isolate their interventions from those of other actors in the ecosystem. This chapter shows that rather than trying to solve the attribution problem, managers are better served by building an *orchestration* system: developing shared (supraorganizational) performance indicators to chart the progress of the field as a whole, support collective learning, and align individual efforts that aim towards joint goals rather than independent aims. I examine the case of Miriam's Kitchen, an organization that aims to end chronic homelessness in Washington, DC, by fundamentally reconfiguring and aligning efforts of over one hundred nonprofit and public sector actors in the city. In concluding this chapter, I draw on the many developments related to a growing "collective impact" movement in the social sector including, for instance, a major strategic shift undertaken by the United Way of America.

These empirical chapters provide a basis for developing a general framework for designing performance measurement and management systems for social change, which I summarize in Chapter 6. They also offer many insights on the roles of funders, which I consider in Chapter 7, in supporting the building of performance measurement and management systems.