



**NEW  
TOOLS TO  
AMPLIFY  
IMPACT**

*A Pay for Success  
Guide to Building  
Nonprofit Capacity*

## ***Authors***

### **JAKE SEGAL**

*Director of Advisory Services, Social Finance*

### **JEFF SHUMWAY**

*Vice President of Advisory Services, Social Finance*

### **JON-PAUL BIANCHI**

*Program Officer, W.K. Kellogg Foundation*

## ***Acknowledgements***

This research was generously supported by the W.K. Kellogg Foundation. For the core content and inspiration behind the work, we are grateful to the four nonprofits who participated in the PFS due diligence process over the past 18 months: AVANCE, Center for Urban Families, the Family Independence Initiative, and HIPPY USA. We specifically appreciate the assistance of the reviewers within each of these organizations: Jesus Gerena (FII), Tenille Patterson (CFUF), Lia Lent (HIPPY), Caitlin Solis (AVANCE). Thank you, as well, to Alex Jutca, Anna Fogel, Rachel Levy, Alexandra Zaroulis, Annie Dear and Kayleigh Burgess at Social Finance, who provided insight and support for this project.

# TABLE OF CONTENTS

## 4 PREFACE

## 6 USING PAY FOR SUCCESS DUE DILIGENCE TO BUILD NONPROFIT CAPACITY

- 11 What we can learn from high-quality nonprofits
  - 12 *Meet the grantees*
  - 15 *Meeting the challenge of measurement*
  - 16 *Using the data right*
  - 18 *Using the right data*
  - 20 *Ensuring data quality*
  - 20 *Moving to an outcomes-based approach*
  - 23 *Communicating value*
  - 24 *Building evidence*
  - 30 *Balancing fidelity with flexibility*
- 35 Putting the tools to work
  - 35 *Nonprofits*
  - 36 *Funders*
  - 37 *Governments*
- 39 A stronger future for the social sector

## 40 A HOW-TO GUIDE FOR CONDUCTING PAY FOR SUCCESS FEASIBILITY ASSESSMENTS

- 42 The purpose of Pay for Success
- 45 Sections of this How-To Guide
- 46 Landscape assessment
- 50 Feasibility assessment
  - 52 *Evidence*
  - 57 *Organizational capacity*
  - 60 *Understanding value*
  - 63 *Identifying payors*
- 65 Improving performance to improve lives

# PREFACE

---

Talk to almost any nonprofit and you'll hear the same story: the organization is making tremendous impact, but financially, it's barely scraping by. Nonprofit leaders have two jobs: achieving their stated mission, and raising money to support that mission. While foundations have long played a critical role as early funders of programs, sustainable funding at scale often comes only with government contracts.

At the same time, governments are stretched. Many feel renewed pressure to get more value from less funding. Government agencies are asking nonprofits to use evidence-based practices and to more carefully measure their outcomes—and they are increasingly tying funding to both. As this movement gathers momentum, nonprofits have begun to feel new strains.

Similarly, for foundations seeking to address social challenges at significant scale, these questions of measurement, growth, and sustainability are critical. Most foundations do not intend to support a given program in perpetuity. At the end of every successful grant lives a question—what next?

This report is an attempt to shed new light on one of the most difficult challenges in the social sector: achieving sustainable funding at meaningful scale. Working with the W. K. Kellogg Foundation and four of their grantees, we explored the question: how can nonprofits prepare themselves—and their data, evidence, and outcomes—for the next generation of government funding?

At Social Finance, we believe that far more public resources should flow toward evidence-based, results-driven solutions—and that such resources should be invested at scale, driving meaningful results across communities. As foundations and nonprofits adjust to this new environment, governments will gain better results for each dollar spent, and organizations with reliably strong results will gain sustainable funding.

Making the shift to results-oriented funding will require new investments for nonprofits: investments in program consistency and quality, data systems, evaluations, and performance management. Increasingly, only those organizations with the capacity to credibly articulate their social and economic value to government decision-makers will receive meaningful government funding.

The purpose of this report is to explore how four nonprofits, all funded by the W.K. Kellogg Foundation, approach this coming shift, and to apply the principles of Pay for Success to each in order to help them prepare for growth via outcomes-oriented government funding streams.







**USING  
PAY FOR  
SUCCESS DUE  
DILIGENCE  
TO BUILD  
NONPROFIT  
CAPACITY**

**IN 2014, THE W. K. KELLOGG FOUNDATION (WKKF) EMBARKED** on an experiment. Like other major funders, WKKF carefully tracked the early workings of U.S. Pay for Success (PFS) projects. In doing so, program officers at the Foundation noticed that PFS was not a single tool, but rather an aggregation of a number of tools. It combined a critical eye toward evidence with an investor lens toward nonprofit capacity, and it produced a hard-nosed estimation of impact on government budgets and on community benefit.

Rather than pursue PFS financing in its entirety, they began to wonder: could the tools being developed for use in Pay for Success due diligence be used independently? More specifically, could they be used to help nonprofits build capacity to track and measure outcomes, and to become better advocates for themselves in conversations with the public sector?

Over the past 18 months, Social Finance has worked in partnership with WKKF and four of its grantees to test this idea. In each instance, we have found that the thinking and tools underlying Pay for Success can be helpful in supporting grantees to clearly define the value they create for participants and for society—building their capacity and improving their ability to secure sustainable public funding.

The purpose of this paper is twofold:

- First, to summarize lessons we've learned from our work with four high-quality WKKF grantees—and to present what others can learn from their efforts; and
- Second, to act as a how-to guide for nonprofits attempting to use the emerging tools of Pay for Success to assess their own capacity and craft a compelling social and economic case for their programs.

We hope that this work can be helpful to nonprofits considering PFS and performance-based contracts and to funders as they consider new ways to support their grantees and spark novel conversations with government.

## WHAT IS PAY FOR SUCCESS?

Pay for Success is about measurably improving the lives of people most in need by driving resources toward better, more effective programs.

At its core, PFS is a public-private partnership that expands funding for high-quality social services through a performance-based contract. PFS projects enable federal, state, and municipal governments to partner with high-performing service providers by tapping private investments to cover the up-front costs of programs. If, following a third-party evaluation, the program is successful in reaching pre-determined outcomes, then government repays the original investment. If the program exceeds those outcomes, the government pays a small return on the investment. If the program does not achieve its target results, government pays nothing.

In this way, PFS ensures that taxpayer dollars are being spent only on programs that actually work—expanding access to quality services for those who need them the most.



# WHAT WE CAN LEARN FROM HIGH-QUALITY NONPROFITS

---

## **IT'S A THURSDAY NIGHT AT A PUBLIC HOUSING PROJECT IN**

San Antonio's West Side, and a dozen young parents are sitting together making toys at the Cortez Family Center. They've come for a weekly, three-hour class on parenting skills, taught by a community member. A few doors down, their young children take part in effective educational programs.

The program, AVANCE's Parent-Child Educational Program (PCEP), is part of a growing two-generation ("2Gen") movement in early childhood development. Like others in the field, 2Gen programs understand the extraordinary value of early intervention in changing the trajectory of a child's life—but also recognize that parents are the most powerful force in teaching and guiding their children. For four decades, AVANCE has been working to build the skills, resilience, and strength of vulnerable families across Texas. As a result, PCEP has taken hold in communities across Texas and is growing nationally. It impacts the lives of thousands of families every year.

A short drive away from Cortez, at AVANCE's national headquarters, the staff is thinking through exactly how to measure that impact on families. Evidence suggests that PCEP improves mother-child teaching interactions, child school readiness, and academic outcomes, and it helps parents connect to the community and find jobs. The impact of the program is both wide and deep. But how much does it improve these outcomes? And how can we track progress of these outcomes over time?

These questions are increasingly important as AVANCE considers growth. Growth requires access to new funding, and maintaining that funding over time. At headquarters, staff members ask: how can we build the case for sustainable funding? How can local chapters partner more closely with—and, ultimately, receive sustainable funding from—the public sector?

# MEET THE GRANTEES



**FOUNDED:** 1973 in San Antonio, TX

**SERVICES PROVIDED IN:** Chapters in Texas; regional office in New Mexico; licensees in California

**TARGET POPULATION:** Primarily Hispanic low-income parents and their children

**MISSION:** “Unlock America’s potential by strengthening families in at-risk communities through effective parent education and support programs”

**HOW THEY DO IT:** The Parent-Child Education Program (PCEP) combines culturally appropriate parenting education, empowerment, and community building with early childhood development services for hard-to-reach families



**FOUNDED:** 1999 in Baltimore, MD

**SERVICES PROVIDED IN:** Baltimore, MD

**TARGET POPULATION:** Young men who are unemployed and/or disconnected from their families

**MISSION:** “Strengthen urban communities by helping fathers and families achieve stability and economic success”

**HOW THEY DO IT:** The Family Stability and Economic Success (FSES) model activates Life Coaches who support CFUF members and can refer them to internal programs targeting chronic unemployment and family instability as well as external services like legal aid or mental health



**HIPPYUSA®**

Home Instruction  
for Parents  
of Preschool  
Youngsters

**FOUNDED:** 1988 in New York; headquarters in Little Rock, AR; program originally developed in Israel in 1969

**SERVICES PROVIDED IN:** 139 sites across 23 states & D.C.

**TARGET POPULATION:** Children and families at risk due to poverty, parents' limited education, or social isolation

**MISSION:** "Partners with parents to prepare their children for success in school"

**HOW THEY DO IT:** A developmentally appropriate curriculum with role play as the method of teaching, with home visits interspersed with group meetings; staffed by home visitors from the community, supervised by a professional coordinator



Family  
Independence  
Initiative

**FOUNDED:** 2001 in Oakland, CA

**SERVICES PROVIDED IN:** Oakland, Boston, Detroit, Fresno, New Orleans, and San Francisco

**TARGET POPULATION:** The "stable poor"—those in or near poverty who are not in crisis

**MISSION:** "Enable America to see the power and potential of low-income families, and enable them to come together and access the resources they need to thrive"

**HOW THEY DO IT:** Arrange group meetings between 6-8 low-income families and enable them to create their own financial goals, solutions, and sense of accountability

Forces reshaping the social sector today lend urgency to these questions. Funders in both the public and private sector, driven by the rise of evidence-based grant- and policymaking, demand ever-greater proof of programs' impact. At the same time, governments around the country are seeking ways to better serve their constituents while spending less money.

As we work with nonprofits to develop Pay for Success strategies, a set of common challenges has emerged as organizations struggle to measure their performance, evaluate their impact, and communicate their value.

These challenges are, of course, intimately—and increasingly—linked. The best organizations use their data to improve performance and achieve better outcomes, and then use evidence of those outcomes to cultivate greater funding. But doing so is hard. In our work with four impressive nonprofits, we saw the true magnitude of these challenges—and how some organizations are striving to overcome them.

## NOTES ON TERMINOLOGY

**Measurement** is the process of tracking data; **performance measurement**, predictably, is when those data describe an organization's performance.

**Performance management** is about using those data to actively improve how the organization operates.

**Outcomes** are measurable results associated with the organization's programs. When a program's outcomes are held up against outcomes from a comparison group, we can determine the organization's **impact**, or the attributable value to individuals and/or society created by a program.

**Evaluation** is the formal process by which many organizations estimate their impact.

## Meeting the challenge of measurement

For nonprofits, mission is what matters. But as any nonprofit leader knows, it can be hard to determine just how well an organization is performing against that mission.

Measurement is where aspiration meets social science. And for nonprofits hoping to influence public policy, impact measurement is the core proof point underpinning the case for change.

But it can be deceptive, too. It's impossible to know exactly what impact a program makes, because we cannot know exactly what would have happened without the intervention.<sup>1</sup> Careful evaluation—which we will discuss in more depth later—can help build a clearer sense of the effect that a program has on society. Strong evaluations are often time-consuming and expensive affairs. In the interim, by collecting and using the right data, programs can build more accurate knowledge of their effectiveness, and can use that knowledge to improve their performance.

Much has been written about nonprofit performance measurement, particularly by consultants such as The Bridgespan Group, McKinsey, and others.<sup>2,3,4</sup> Good performance measurement is grounded in and tailored to the organization's theory of change; it's disseminated through the organization's culture, but led by dedicated experts; and it's purpose-built to help people make decisions. Measurement solely to appease funders becomes anemic, while measurement used for management becomes essential.

Really strong performance measurement, though, is hard. Program participants rarely lend themselves to easy tracking; they change phone numbers, names, cities. Important outcomes can sometimes

---

1 For the purposes of this paper, we typically refer to interventions—both the generic models of social change, as well as specific organizations' versions of those interventions—as “programs.” We continue to use the term “intervention” when it refers to the action or process of intervening (“improving the case for intervention”), rather than describing a model of social change.

2 Performance Measurement. Bridgespan Group, 2015. Web. October 2015. <http://www.bridgespan.org/Publications-and-Tools/Performance-Measurement.aspx#VmXAG3arTIU>.

3 Measuring What Matters in Nonprofits. McKinsey & Company, 2015. Web. October 2015. [http://www.mckinsey.com/insights/social\\_sector/measuring\\_what\\_matters\\_in\\_nonprofits](http://www.mckinsey.com/insights/social_sector/measuring_what_matters_in_nonprofits).

4 See, e.g., “The New Measures for Improving Nonprofit Performance.” HBS Working Knowledge, 14 Dec. 2011. Web. October. 2015. <http://www.effectivephilanthropy.org/portfolio-items/room-for-improvement>.



Measurement solely to appease funders becomes anemic, while measurement used for management becomes essential.

take years to occur. Essential data—particularly administrative data collected by government departments for their own records—are often closely protected and challenging to access.

Nonprofit leaders must build the skills and cultivate the organizational behaviors required to put these data to good use. In our work with AVANCE, Family Independence Initiative (FII), Home Instruction for Parents of Preschool Youngsters (HIPPI), and the Center for Urban Families (CFUF), we saw the challenges of performance measurement come to life—and pathways to overcoming those challenges.

### ***Using the data right***

For years, one organization we've worked with collected a handful of paper questionnaires, asked volunteers and staff members to transcribe them onto machine-readable Scantron sheets, packed them into crates, and shipped the crates to their headquarters to be entered into a database. The crates sat in a corner of a crowded office until the month before the organization wrote its annual report. One year, a major site's data was lost to a basement flood before it could be recorded.

Getting data from the street to the spreadsheet can be plagued by challenges. Paper, though, is not the real enemy here. It's

access. Data, recorded on paper or otherwise, are too often stored in inaccessible archives. Overtaxed analysts spend inordinate amounts of time building and refining reports. Another organization we've worked with collects exquisitely detailed operational information, but because of how hard their database is to work with, has developed a policy that requires senior-level signoff for all custom reports. Using cumbersome data for ongoing analysis and decision making—for answering questions as they arise—is out of the question.

But some nonprofits are taking data collection and analysis to a new level. Take, for example, the Family Independence Initiative (FII), an Oakland-based group pioneering a novel approach to supporting economic and social mobility in America. FII arranges group meetings of between six and eight low-income women who create their own financial goals, solutions, and sense of accountability. Like any good team, the group learns to turn to one another for motivation, support, and advice, while FII's resource bank of financial products aids them along their path to self-sufficiency.

At FII, monthly journal entries are uploaded into a sophisticated data collection system through a cloud-based, real-time data management platform, InvestCloud. Progress is quantified and tracked on both quantitative metrics (such as income) and subjective measures (such as well-being). The dataset allows FII to confirm whether, for example, changes in family income for a group of participants were driven by an increase in public benefits or by an increase in employment income. High-quality data collection allows FII to refine its program and to quantify its results for external stakeholders.

Most organizations, though, aren't FII. Systematic underinvestment in technology hamstrings otherwise great organizations.<sup>5</sup> Nonprofits underestimate the resources required to buy, integrate, and maintain technology systems. Funders demand extensive data, but they are rarely willing to pay for extensive data collection.

At the same time, technology solutions are not solutions in their own right. They are tools. For good data capture tools to be useful, they need to capture the right data—data that are both useful and accurate.

---

5 See, for example: Anna Goggins Gregory and Don Howard. "The Nonprofit Starvation Cycle." Stanford Social Innovation Review, 2009. Web. Oct. 2015. [http://ssir.org/articles/entry/the\\_nonprofit\\_starvation\\_cycle](http://ssir.org/articles/entry/the_nonprofit_starvation_cycle).

## Using the right data

When data systems aren't kept on a tight leash, they can become tyrants. When thinking about measurement, organizations often ask, "how should we measure?" when they should be asking, "what should we measure—and why?"<sup>6</sup>

Leadership at HIPYPY took these questions to heart. HIPYPY supports low-income parents of young children through home visits and group meetings, in which parents learn how to help their children become ready to succeed in school. Tracking information as parents move through the program is essential to understand and evaluate program impact. HIPYPY's home visitors are asked to enter key data points after every visit. Over the course of our work together, HIPYPY's staff and leadership reviewed and refined their tracking until they could identify and agree on the value of each data point collected. In parallel, staff worked out a way to share the burden of measurement: home visitors now gather data, while the deputy program director enters it.

At FII, data are essential to the program not just for frontline staff, but for participating families themselves. The program has defined a set of key questions that, in their own right, help to support families toward sustainability. As participants enter data on actions taken and outcomes achieved over the last month, they are creating a personalized history of progression towards their own goals. The data are always available to them online, and serve as a benchmark and motivation for families through the FII journey, just like an exercise-tracking app on your phone.

Simultaneously, the data provide FII with real-time information about where families are succeeding or struggling, prompting changes to the resources provided in the FII Resource Bank—a unique set of financial and other supports tailored to the needs of FII participants. For instance, FII launched lending circles through a partner, Mission Asset Fund, to provide financing to FII families and help them rebuild credit scores, serving two needs highlighted through data collection.

---

6 Morino, Mario. *Leap of Reason: Managing to Outcomes in an Era of Scarcity*, Washington, DC: Venture Philanthropy Partners, 2011.



## **Ensuring data quality**

Data are never perfect. But major data gaps and entry errors can lead organizations to misleading conclusions. Intermittent reporting can exclude the most important data; inaccurate data can skew results. These are challenges that can plague even organizations with advanced data collection systems such as FII.

But these challenges also present opportunity. Thoughtful approaches to the entry and audit process can enhance the quality of data. For example, in our work together, FII and Social Finance worked to proactively protect against inaccuracy. In our review of FII's family data, we found that self-reported responses sometimes (unsurprisingly) contained input errors. In response, we jointly built into the collection tool a series of automated checks. Potentially inaccurate entries are now immediately noted for families and FII liaisons at the time of data entry; FII liaisons can then follow up with families to confirm during their next meeting whether, for example, their monthly income is three standard deviations above that of the average program family. In situations with straightforward solutions—misspelled words, or numeric variables with an incorrect sign (e.g., negative personal income)—the new tool automatically fixes the mistake and standardizes the answer format.

Ongoing monitoring of data quality is an important part of good performance measurement. But tools alone are not enough; organizations need to be strategic and intentional about how they use them.

## **Moving to an outcomes-based approach**

Outputs count what gets done; outcomes describe why those outputs matter. Even tracking the easy stuff—names, dates, places—isn't so very easy. Tracking outcomes is tougher still. Only rarely do nonprofits track participants for long enough to determine outcomes, and even more rarely are they judged against a rigorous comparison group.

The first step is to define which outcomes a program intends to influence. Defining an organization's key outcomes is a plank in good nonprofit strategy. Outcomes describe the kind of impact an



Getting data from the street to the spreadsheet can be plagued by challenges. Paper, though, is not the real enemy here. It's access.

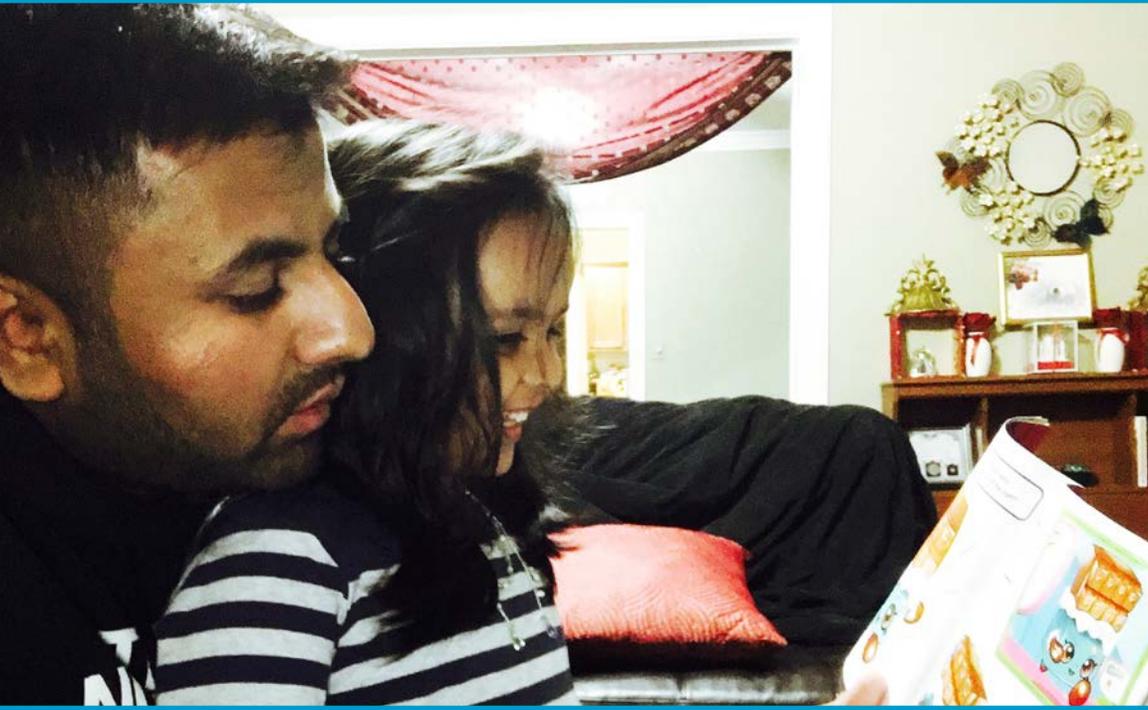
organization will hold itself accountable for—one part of what The Bridgespan Group calls its “intended impact.”<sup>7</sup>

But at the same time, a concise statement of positive program outcomes can be hard to pin down. After all, many good programs create diffuse impact that reaches many different participants in different ways.

Take AVANCE, for example. For over four decades, the Texas-based nonprofit has been a centerpiece of the growing 2Gen movement, working with both parents and children to improve family engagement and early childhood development. PCEP stretches over nine months and a broad range of topics. True to its philosophy, AVANCE believes that its work creates outcomes for parents—such as increasing employment, improving mental health, and boosting educational achievement—as well as children—bolstering academic achievement, health, and welfare.

---

7 The Bridgespan Group defines intended impact as: “...a statement or series of statements about what specifically the organization is trying to achieve and will hold itself accountable to. It succinctly identifies what results the organization will accomplish, for whom, and in what time frame. While these frequently serve as complements to mission statements that tend to be fairly aspirational in character, they may actually be the same as mission statements where expected outcomes are set forth with sufficient clarity.” *Intended Impact/Theory of Change*. The Bridgespan Group. Web. 2015 Nov. <http://www.bridgespan.org/Publications-and-Tools/Nonprofit-Management-Tools-and-Trends/Intended-Impact-Theory-of-Change.aspx>.



PCEP is an evidence-based program. It was studied in a high-quality evaluation funded by the Carnegie Corporation of New York in 1991,<sup>8</sup> which provided strong evidence of improved mother-child teaching interactions. Subsequent qualitative research suggests additional and wide-ranging benefits of the program. These data points, taken alongside AVANCE's strong leadership and deep community engagement, have helped AVANCE successfully raise funds to expand nationally.

Like many of the nation's best nonprofits, AVANCE has been tracking near-term outcomes for years, often using self-reported data. But the outcomes it tracks are not always compelling to public-sector policymakers. Tailoring AVANCE's outcomes to local priorities—for example, demonstrating for budget-constrained school districts not only improvements in student achievement, but also improved attendance, which positively affects a school district's bottom line—can help build support among policymakers. Building grassroots relationships with families and communities

---

8 Carnegie Study. AVANCE, 2013. Web. Oct. 2015. <http://www.avance.org/carnegie-study/>.

is in AVANCE's DNA; as they build partnerships of greater scale, the program is focused more and more on integrating their own data with those of districts themselves and using those data to demonstrate the strength of their outcomes.

## Communicating value

AVANCE is not alone in seeking to clarify the policy-relevant outcomes it influences. In early 2014, CFUF started down a path of organizational and cultural transformation, integrating several programs (employment, fatherhood, couples) under a unified model. The Family Stability and Economic Success (FSES) model targets chronic unemployment and family instability using Life Coaches<sup>9</sup>

to connect participants—usually young men who are unemployed and disconnected from their families—with specific programs, the CFUF community, and outside services. Nevertheless, it is difficult to quantify and communicate the unique impacts of the emerging FSES model. Despite the challenges, programs like FSES can use the tools of Pay for Success to estimate and

communicate the tangible value they create for society. While FSES has not been evaluated at an aggregate level, and does not yet have access to the right administrative data needed to build their own rigorous evaluation, evaluations of distinct components within FSES provided a starting point for understanding the program's aggregate impact. One component, STRIVE Baltimore—a workforce development and employment readiness program for men and women—has demonstrated a long history of positive effects on employment attainment and wages for its graduates; another, the Baltimore Responsible Fatherhood Project (BRFP), a national program intended to increase fathers' emotional and financial support of their children and families, can draw on evaluations of similar programs that suggest positive effects on everything from employment to the relationship between father and child.<sup>10</sup>

### FOR MORE INFORMATION

*See the How-To Guide,  
page 60*

9 Upon entering into FSES, through a process called Intake, a Life Coach is assigned to each participant to act as their guide, mentor, and connection to the organization. Life Coaches can recommend programs based on need, refer participants to external partners, and develop deeper and more meaningful connections with each participant.

10 Process evaluations analyze the effectiveness of program operations, implementation, and service delivery rather than the outcomes of the program. BRFP is currently engaged with the Fatherhood Research & Practice Network to use a randomized controlled trial to test the efficacy of its curriculum.

We worked with CFUF to draw on field research and academic evaluations to estimate the combined value that FSES creates for Baltimore's families. This process admittedly takes a narrow view of CFUF: it quantifies the value of a few measurable outcomes and uses these values as a proxy for the entire program's value. We discuss in more detail in the how-to guide how Social Finance makes recommendations about estimating program value, but the reality is that while programs create far more value than we can readily quantify, putting figures to what we can quantify is a good start. It can provide a compelling, if incomplete, picture of an organization's value to public-sector stakeholders who often rely on the skills and diligence of nonprofit colleagues.

Like CFUF, HIPPOY believed that its program created value for its children and families, both during the program and long beyond, but it wasn't clear how to measure or quantify that value. Leadership spoke of strengthening family ties, increasing the time parents spend with their children, and bolstering readiness to learn in kindergarten. But how could HIPPOY quantify the value of these important changes? And are they only valuable to the families themselves, or are they also worth something to school districts or local governments?

We worked with HIPPOY to identify relevant administrative data and benchmarks for its program in different sites in order to find the right comparison points. There is compelling evidence that children who go through HIPPOY programs will be less expensive to educate. They are prepared to learn in kindergarten and are less likely to require remedial education or special education. A year of special education costs more than a year of regular education; we were able to use this cost differential as a starting point to calculate the value of HIPPOY's program to a school district.

Of course, one of the hardest parts of this exercise is the crucially important outcomes we didn't include in the valuation. Such an exercise reflects a fraction of the value created by the program. But, it begins to translate that portion of value into something palpable, something clearly defined, and sometimes—when it plausibly achieves a goal for government, such as reaching a policy priority or saving money—something genuinely appealing to the public sector.

## ***Building evidence***

As we've seen, organizations like HIPPOY, FII, AVANCE, and CFUF use data systems that they can access in real time; they're thoughtful about the data they collect, and how it will be useful; they invest



time to make their systems useful and accurate; and they describe their impact in terms of outcomes that are relevant to policymakers.

But how do organizations actually measure those outcomes?

**FOR MORE  
INFORMATION**

*See the How-To Guide,  
page 52*

**INVESTING IN EVALUATION**

Of course, one way to measure outcomes is by running experiments. Randomized controlled trials (RCTs) are the gold standard in social science research. And for good reason: they avoid the potential biases that challenge the findings of so many nonprofit evaluations. (For more on biases and other threats to credibility, see “Evidence in Pay for Success feasibility studies” in the how-to guide later on.<sup>11</sup>)

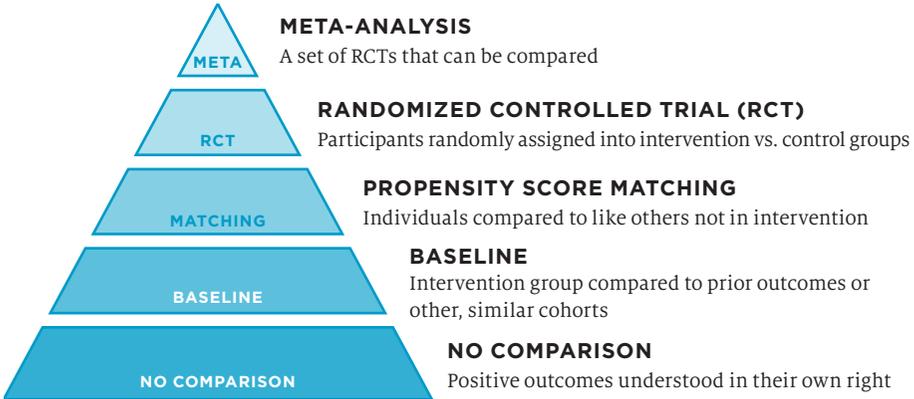
Experiments can provide useful information about a program’s impact. However, RCTs are remarkably rare. They can be expensive, lengthy, and challenging to operationalize. In isolation, they may not be easily generalizable—reflecting instead the specific time, place, and population of the experiment itself. Few funders—either

---

11 Others have written about the value of randomized controlled trials extensively, and have suggested methods for conducting such experiments at lower costs. See, for example, *Demonstrating How Low-Cost Randomized Controlled Trials Can Drive Effective Social Spending: Project Overview and Request for Proposals*, Coalition for Evidence-Based Policy, 2013 Dec. Web. 2015 Oct. <http://coalition4evidence.org/wp-content/uploads/2014/02/Low-cost-RCT-competition-December-2013.pdf>.

**FIGURE 1** Evidence of impact

Based on graphic first published by UK Cabinet Office in the SIB Knowledge Box; sourced from Emma Tomkinson, “What do we know about the Utah SIB results (without a counterfactual)?,” 2015 Nov. Wed. 2015 Nov. <http://emmatomkinson.com>.



philanthropic or governmental—spend the resources to deeply evaluate the success of their investments.

### **ACCESSING ADMINISTRATIVE DATA**

In the absence of experimental data, retrospective studies can be hugely valuable to build nonprofit evidence.

Often, the data needed to understand a program’s outcomes exist and are already tracked by government agencies. These data—what we call administrative data, the data that agencies collect in the course of regular operations, rather than through specific evaluations—can be extraordinarily useful. They can track outcomes not only for participants of a program, but also for others who were not a part of the program. By carefully matching participants to non-participants using statistical techniques, we can make reasonable estimates of the impact of a program versus a comparison group.

But administrative data can be hard to obtain. Data are often housed within agencies that are stretched for capacity and reluctant to share with external researchers. Accessing certain datasets means complying with stringent privacy standards. Even when data are accessible, matching them to participant records held by the nonprofit organization can be challenging.

Nevertheless, administrative data can be a powerful tool for building an investment case. Often, organizations with unique competencies like AVANCE can't simply rely on evaluations of similar programs to describe their impact. Indeed, the early childhood field is starved for high-quality evidence of what works among Hispanic populations. According to one early childhood expert, "there are no longitudinal data whatsoever about how early childhood education impacts Hispanic children." AVANCE is trying to change that. The organization works with the Dallas Independent School District to evaluate PCEP's impact, and has begun planning for an in-depth evaluation of its fatherhood program funded by the Administration for Children and Families. Evidence from these efforts will contribute important findings to the field.

In order to build meaningful evidence, AVANCE has had to think carefully about which outcomes to track. While AVANCE benefits participants in many parts of their lives, the organization can't realistically seek to access administrative data in health, public housing, labor, public benefits usage, educational success, and foster care. Instead, to bolster its case, AVANCE looked to the intersection of where it makes the largest difference and where it could plausibly access data. Similar programs, and AVANCE's own experience, suggested that educational outcomes—from student achievement to grade promotion—are likely the most fruitful for exploration.

Crucially, AVANCE's chapters have been careful to build close relationships with school districts that own these data. In Dallas, for example, the AVANCE chapter worked diligently with the Dallas Independent School District to ensure the District had flagged former AVANCE students in its records. This work has helped the program in its tracking efforts. It also laid the groundwork for a recent internal District evaluation, which suggested positive impact of AVANCE's programs on attendance and school achievement.

Similarly, HIPPIY and Social Finance explored potential avenues for accessing administrative data to describe HIPPIY's impact on its families. In Arkansas, where HIPPIY serves more than 5,000 families across the state, the organization partnered with the Arkansas Research Center (ARC), a nationally renowned state longitudinal data system which gathers data on every student in any Arkansas public school. The Arkansas Research Center had never created a flag in their data to identify students that had been through a HIPPIY program—or, for that matter, any other early childhood program—but they worked closely with HIPPIY to do so,

which presents new opportunities for analysis both at HIPPY and at ARC. This relationship will allow HIPPY to view the academic outcomes—including kindergarten readiness, standardized test scores and special education utilization—for children served by the program.

Building trust with public-sector partners is much like building trust with the community, and nearly as important. These relationships—whether with local school districts, county departments, state agencies, elected officials, or quasi-governmental research centers—are vital when it comes to conducting research on program effectiveness. Few organizations, though, pay them adequate attention.

In our work with AVANCE and HIPPY, and drawing on the experiences of other nonprofits, we developed a set of **best practices in strengthening public-sector relationships**:

**1 CREATE A TWO-WAY PARTNERSHIP**  
Organizations that demand data without developing a true partnership with the people who hold those data typically aren't successful in their efforts. Those that are successful demonstrate their value and the value of the data being disclosed—by sharing back valuable analysis, organizing their volunteer networks to support partner agencies (e.g., school districts), or opening their own internal professional development trainings to relevant public-sector colleagues.

**2 BUILD A NETWORK**  
School district leaders can sometimes feel that early childhood programs like AVANCE and HIPPY are removed from the day-to-day activity of their schools. This can be reinforced by lack of proximity: some of AVANCE's programs are taught in community centers, while HIPPY's services take place in the home. Despite the distance, successful chapters used their touchpoints to deepen relationships. They made a point of telling superintendents and principals—not to mention mayors and city councilors—of their results, and what those results meant for schools. High-level support consistently helped to expedite program integration and deeper partnerships. At the same time, they connected with school staff members working directly with data and with students to build strong ground-level support and to build a spirit of collaboration.

**3 BE CLEAR IN REQUESTS, AND PROVIDE CONTEXT**  
Administrative data, as we've mentioned, can be hard to access. The broader the request for data, we've found, the less likely a partner is to share. Requests should be targeted narrowly, and the



rationale behind them—the programmatic data or the desk research that suggests a given program can affect a given outcome—should be made clear to on-the-ground staff.

#### **4** LEVERAGE EXPERTS

Partnerships—such as those with evaluators or academics—can contribute connections, credibility, and experience, add weight and urgency to discussions of evaluation, and reassure public-sector partners as they work through challenges to accessing data.

We believe that these relationships—and, ultimately, the ability to access the administrative data which these relationships can help to unlock—are essential for high-quality nonprofits to build their evidence and create pathways to sustainable public funding.

### ***Balancing fidelity with flexibility***

The best nonprofits learn as they grow. But to understand the lessons of the past, it's important not to change the program so dramatically that it becomes something new altogether. How much evolution, then, is the right amount for a program as it grows?

Over twenty-five years, HIPPY USA has grown into a national program serving more than 15,000 children in 23 states and the District of Columbia. While the program in rural Arkansas doesn't look exactly like the program in urban Milwaukee, they share key similarities. Every program follows four main guidelines: a developmentally appropriate curriculum; teaching through role play; staffing with a site coordinator and home visitors; and delivering services in the home and in group meetings. HIPPY USA, in partnership with its state offices, certifies that local programs meet these guidelines, ensuring all programs are held to the same standards and are using similar materials and curricula.

But HIPPY's programmatic guidelines leave room for flexibility. Local providers can make adjustments to meet the needs of local families. When appropriate, home visitors can adjust to meet mothers at work, or can slow the cadence of visits for mothers who need fewer touchpoints. These refinements allow families who otherwise might drop out to stay in the program. Adherence to the core model is important in developing a program's evidence, but flexibility within that architecture allows the program to maximize its potential.

## BEING A PART OF THE COMMUNITY

**There are many kinds of fidelity. While we often focus on fidelity to a program model, fidelity to the community is similarly essential.**

Cultural competency lives in the DNA of an organization. For each of the WKKF grantees, creating deep and honest connections with the communities they serve has been critical to their success.

**CFUF** is in a part of Baltimore called Greater Mondawmin—a historically African-American community and one of the most diverse areas in the city, largely blighted, just steps away from where Freddie Gray was arrested in April 2015. Creating a safe haven in the heart of this community was a strategic choice for CFUF. Nearly one-fifth of the organization's participants simply walk in from the surrounding neighborhood. Another third come from direct outreach by staff into the nearby community: CFUF employs an outreach team to recruit potential clients city-wide and provides a stipend to program graduates to assist with recruitment in the neighborhoods where they reside. Its Mondawmin presence lends CFUF both credibility with and access to the residents it is trying to reach.

**AVANCE**, for its part, builds this connection starting with recruitment. Staff walk door-to-door to speak with potential participants in their homes. After its participants go through the program, AVANCE actively converts them into parent educators, ensuring the program stays closely connected to and embedded in communities. According to estimates from the organization's national office, over 45% of AVANCE's 900 staff are former graduates of the program.

Creating this community connection has been an essential ingredient to success. But it is also an important challenge to scale. As AVANCE enters new communities, it has to rethink this relationship with the community. When setting up new sites, building these relationships can take years. Partnering with trusted local organizations can accelerate that process. But for AVANCE, as with all organizations we've worked with, understanding and codifying the "secret sauce" of community connection—and making smart choices about where and how to expand beyond its current footprint—can be a struggle. The tradeoffs between spreading AVANCE's methods to new geographies and expanding AVANCE's reach within its current sites is a significant point of tension in the organization's growth strategy.

As nonprofits grow, many face these tradeoffs. Often, Pay for Success-funded expansion into new geographies can be riskier than deepening services within current sites. Community standing and rapport are, for many programs, essential to success. Geographic expansion asks providers to assess opportunities and partnerships in new ways, or even to build referral pipelines and community relationships from scratch.



We get to see the data, and perhaps most importantly they get to see their data. Many families use the data as a tool to keep track of their own progress. And we ask them to help us understand the stories behind the data. Not only have we seen families do significantly better, but we've seen ripple effects in communities, where people hear about these successes and they are inspired to do more too.

**MAURICE LIM MILLER, FOUNDER  
AND CEO, FAMILY INDEPENDENCE  
INITIATIVE (FII)**

*From Transforming Latino Social Mobility: An Interview with Maurice Lim Miller in The Harvard Journal of Hispanic Policy, in reference to FII's unique data collection system*



# PUTTING THE TOOLS TO WORK

**THE LESSONS WE'VE LEARNED FROM WORKING WITH WKKF** grantees and from applying Social Finance's feasibility assessment framework suggest the need for action among all stakeholders involved in Pay for Success.



## *Nonprofits*

Trends in the industry will continue to promote the growth of organizations dedicated to collecting and using data. Increasingly, government and funders will demand a robust and proven strategy for how nonprofit services will directly lead to outcomes.

Nonprofits **should continue to build out their performance measurement systems**, carefully deciding which data to track, how to track them, and how to use those data in real time to inform their work. They should link those data, ultimately, to outcomes that are relevant to the priorities of public-sector funders, and should seek to evaluate those outcomes against carefully constructed comparison groups in order to understand their program's true impact.

In service of those goals, nonprofits should make it a key priority to **forge deep, two-way partnerships** with government agencies that hold administrative data.

They should also work with their Boards and their funders to **strengthen how they communicate their value proposition**. Good nonprofit strategy has for years been founded on clear statements of organizational theories of change and intended impact. Pay for Success demands more. It asks organizations not only to hold themselves accountable to outcomes, but to define the value that those outcomes create for society. Even for organizations that are not pursuing Pay for Success, this kind of analysis can be effective. It demonstrates clearly the fiscal and community benefit of a given program, building a compelling funding case and a rationale for weaving that program into the fabric of public-sector procurement.



## Funders

Funders are demanding more of grantees—more evidence building, more data, greater focus on outcomes. They also need to give nonprofits the support required to meet those demands.

While funders should continue to underwrite innovation, they should place greater emphasis on—and begin to truly incentivize—programs that have invested their time and effort into building evidence of impact. Funders, through their grant making, influence the priorities of nonprofits; as the market for outcomes develops, funders can support and scale those organizations best positioned to succeed and build their capabilities.

Funders should invest more deeply in these grantees. Building data systems, linking them to outcomes, and developing formal evaluations are challenging and expensive. Without a funder acting as champion, most nonprofits can't afford these activities, which are essential to building evidence. It is not enough to prioritize outcomes and evidence; funders must support grantees with **bigger, more concentrated grants** to do better research and build the strategic infrastructure and relevant capacity essential to delivering and measuring high-quality results.

Funders should **dedicate resources to research** by supporting access to administrative data, third-party program evaluations, and internal evaluation expertise. They should help to **build performance measurement capacity** by enhancing the ability to collect, share, and analyze data in order to drive decision-making. And they should support **development of a next-generation business case** by helping nonprofits to do the hard work of quantifying their value—defining the outcomes that they influence and estimating the fiscal and social value of those outcomes to key funders. Demonstrations and pilot projects can further strengthen the case for impact on specific outcomes.



## Governments

Governments play a central role in driving towards a stronger social sector. Procurements at the local, state, and federal levels fund billions of dollars' worth of social and human services each year.

Innovators within government can take the lead in building a performance-based, outcomes-driven sector. They should start by asking the hard questions about the outcomes they're seeking. How much is it worth, from both a fiscal and community benefit perspective, to help a child avoid the need for special education? What is the value of preventing diabetes among low-income seniors? How should we measure the impact of housing the chronically homeless? How, then, do we assess the long-term benefits generated for individuals and our community of achieving these shorter-term outcomes?

**A more robust understanding of the true costs and benefits of key societal challenges can unlock a new world for government procurement.** It can allow the public sector to leverage smart funding mechanisms that achieve the greatest value for money—including Pay for Success contracts, where appropriate, but also other kinds of performance-based contracts (paying providers based on the outcomes they achieve) and even just performance-based budgeting (choosing providers based on a deep understanding of their evidence).

At the same time, governments play an essential role in **facilitating evidence building**. If the public sector is truly committed to improving outcomes, data sharing is not a luxury—it's a necessity. Nonprofits need better access to administrative data to measure their value and to course-correct when they underachieve.



# A STRONGER FUTURE FOR THE SOCIAL SECTOR

---

## **THE SOCIAL SECTOR HAS CHANGED DRAMATICALLY OVER**

the last twenty years. Increasingly, funders and nonprofits are focused on collecting data, analyzing it, and matching it against reasonable comparators.

Pay for Success builds on this movement. It takes the evaluations and the program data compiled during an era of focused philanthropy and interrogates them in an effort to estimate the value created by the underlying social sector programs. In doing so, we hope to make explicit the usually implicit tradeoffs that drive public-sector procurement—spurring a conversation that focuses on outcomes and on value—and ultimately change the way that governments buy social services.

This has changed the relationship between nonprofits, funders, and their public-sector counterparts. Nonprofits are beginning to use their data not just to improve performance, but to make a case for sustained funding and expansion. Governments are beginning to consider the estimated value of programs and the likelihood of their success to inform funding decisions. Great nonprofit leaders and community voices remain as important as ever; but now, they're being complemented by thoughtful analysis of a program's fiscal and community benefit.

In this context, nonprofits must better advocate for themselves. They, and their funders, should continue the thoughtful work of nonprofit strategy and business planning that blossomed at the turn of the century, and at the same time perform analyses that more fully link their data to outcomes in order to make their case to the public sector.

In the next section, we discuss methods for conducting self-assessments using the lens of Pay for Success—with an eye toward supporting leaders as they make the case for large-scale government investment.





*Moving Toward  
Performance-Based  
Financing:*

**A HOW-TO  
GUIDE FOR  
CONDUCTING  
PAY FOR  
SUCCESS  
FEASIBILITY  
ASSESSMENTS**



# THE PURPOSE OF PAY FOR SUCCESS

---

**PFS OFFERS A WAY TO CREATE A MORE OUTCOMES-** oriented social sector. While we are committed to mobilizing capital to scale up effective solutions in individual PFS projects, these project are, in many ways, not the ultimate goal of PFS. Success is about changing the way governments think about contracting for services. Today, governments typically buy services or outputs. PFS allows government to instead buy outcomes—and to know through rigorous measurement exactly what they are getting.

The transformational power of the tool is in changing the way governments allocate funding. Governments collectively spend hundreds of billions on social services every year, often based on limited evidence and with little or no measurement of results. Pay for Success helps bring governments to the table to make decisions based on evidence and pay only for results. At the heart of it, Pay for Success is simply a financing tool to help reorient government spending towards better results in our communities.

We envision a future in which cities, counties, and states routinely perform detailed cost-effectiveness analyses on potential human services investments, rigorously review provider evidence, fund programs at scale based on their expected impact, and support those programs with ongoing measurement and performance management.

At the same time, we're mindful that the PFS model is not appropriate in many instances. Every day at Social Finance we speak with great nonprofits working to solve the puzzle of intergenerational poverty. Only a small fraction of these organizations, though, are operating in the right conditions for Pay for Success.<sup>12</sup>

---

<sup>12</sup> More specifically: Pay for Success as it is constituted today in the United States. For many reasons, the market at this writing emphasizes interventions with extraordinary evidence and a strong track record. Many deals have returns that are artificially capped by governments; above a certain level of upside,

Nevertheless, our work with WKKF and its grantees suggests that even if Pay for Success is not right for an organization, many of the tools and assessments used in Pay for Success can be helpful to nonprofits as they think about their own strategy and communications.

It is with these nonprofits in mind that we describe our PFS due diligence process in the following sections. For many organizations, the kinds of evidence and capacity required for PFS may not be right; but at the same time, the kind of thinking and analysis underlying this process may be valuable as they measure and demonstrate their value.

---

payors receive all of the benefit, and investors none. Many impact investors interested in PFS focus on return of principal rather than maximizing risk-adjusted returns. The “venture capital” model of PFS—in which investors put their money against true innovation, rather than simply taking on the (often considerable) performance risk of scaling up proven practices—is rare in the US. As the market evolves, however, these circumstances may change, which would influence the guidance in this paper.

# SECTIONS OF THIS HOW-TO GUIDE

---

To assess readiness for PFS within a geography, we typically apply a two-step process.

## LANDSCAPE ASSESSMENT

Through an initial screening process which we call a **landscape assessment**, we seek to identify community needs and potential solutions, and assess their fit with PFS. This work yields an initial set of high-quality programs and potential PFS opportunities. This may mirror, or even challenge, the process that thoughtful funders undertake in their grant making. For most nonprofits, this section may be less relevant, so we've kept it brief.

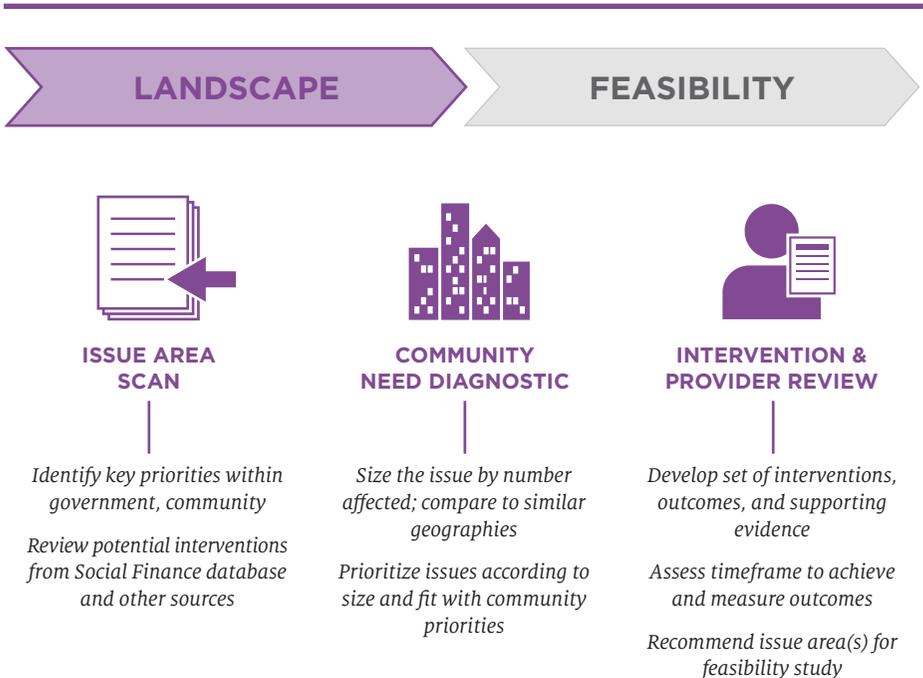
## FEASIBILITY ASSESSMENT

In a second, deeper analysis that we call a feasibility assessment, we dive into each element of the program's evidence and the provider's capacity and apply economic analyses to determine if there is sufficient value for society to support a PFS financing. This section may be a useful guide both for nonprofits and for funders in both the public and private sectors.

# LANDSCAPE ASSESSMENT

At Social Finance, we work with funders and governments to assess a wide range of programs and providers for their PFS readiness. Often, the work begins when **local leaders identify a set of key challenges** facing their community, such as high rates of preterm birth, recidivism, or unemployment. As a first step toward solving those challenges, funders or governments ask us to assess which challenges can plausibly be prevented using Pay for Success within their jurisdiction, considering both the needs of the community and the ecosystem of high-quality local and national programs.

**FIGURE 2** Landscape study overview



The first stage of this work involves working with stakeholders in the community to identify top-priority challenges with an **issue area scan**. In support of this work, we draw on local data about the magnitude of these challenges—better characterizing their size, geographic and demographic distribution, and trends affecting their future prevalence—as well as on interviews with civic and community leaders, existing community needs surveys, and stated local policy priorities. Finally, we identify high-quality interventions that can help to tangibly influence key outcomes within each issue area and make recommendations to more deeply explore a single issue and a narrower set of interventions—ideally just one—in more depth.

Central to this work is **identifying and assessing relevant programs** within a given issue area, in order to narrow the universe of potential programs before conducting deeper diligence. In the course of doing so, we often look at dozens or even hundreds of programs and the nonprofits that deliver them, assessing their fit for PFS.

In our initial explorations, as summarized on the next page in Figure 3, we typically use just a few key criteria to determine that fit.

**FOR FUNDERS, these criteria may be helpful to think through in your own strategic landscaping work; FOR NONPROFITS, they may be helpful in understanding how ready your organization is for performance-based funding.** In our deeper diligence process, as we'll describe further in the next section, we delve into these attributes in significantly greater detail.

## DETERMINING FIT FOR PAY FOR SUCCESS

<b>DEFINED TARGET POPULATION</b>	<i>Underserved, large-scale population with adequate demand for intervention</i>
<b>CODIFIED PROGRAM MODEL</b>	<i>Well-codified program model with fidelity monitoring</i>
<b>RIGOROUSLY EVALUATED</b>	<i>Evidence base studied versus rigorous comparison group</i>
<b>SCALABLE SERVICE PROVIDER</b>	<i>High-quality provider with capacity to scale</i>
<b>POSITIVE ROI</b>	<i>Clear link to public-sector benefits (economic and community benefits) within reasonable timeframe</i>
<b>PAYOR AND INVESTOR INTEREST</b>	<i>Outcomes attract civic and/or commercial support</i>

**FIGURE 3** Screening criteria for programs pursuing PFS financing

### DEFINED TARGET POPULATION

We look first to understand beneficiaries. What do we know about which populations benefit most from the program? How well will the impact of a program translate to the target population among which it would be replicated? This includes not only looking at standard demographics (age, race, gender), but also local context (including community resources) and individual risk factors (e.g., prior health conditions).

### CODIFIED PROGRAM MODEL

Next, we try to understand the replicability of that impact. To do so, we need to look at how well defined the program is: does the nonprofit adhere to a standard, codified framework of service delivery in which certain program elements are flexible and adaptable as appropriate? Is a system in place to monitor adherence to the model? If the answer to these questions is yes, then it's more likely that its previous impact could be replicated.

### **RIGOROUSLY EVALUATED**

Evidence is at the heart of the Pay for Success model—and it's also typically the greatest hurdle for programs. We typically focus on the existence of high-quality evaluations that seek to establish a causal link between a program and the outcomes we are seeking. Specifically, we are looking at the quality, relevance, and findings of these evaluations, in terms of the percentage change generated on outcomes that public payors are likely to care about. We also look at evaluations of similar programs and an organization's internal programmatic data as a key part of the process.

### **SCALABLE SERVICE PROVIDERS**

Next, we turn to the providers themselves that offer (or could offer) the program, seeking to understand their track record and their organizational strengths and weaknesses. This includes the strength of the organization not just in terms of its finances or operations, but also its ability to use data to track and improve programming and outcomes, and its connection to the community it serves. We delve much more deeply into these questions during the Feasibility phase.

### **POSITIVE ECONOMICS**

To be a good fit for PFS, the program must have a clear value proposition, creating benefits—for both the public sector and society—that outweigh the costs of providing the services. Again, during the initial evaluation, we look for indicators of positive economics; in the Feasibility phase, we do a more intensive cost-benefit analysis. We discuss this in more detail in the “Understanding value” section.

### **INTEREST FROM PAYORS AND INVESTORS**

Finally, and crucially, the outcomes produced by the program must align with the policy priorities of a payor. A government or a private payor (such as a managed care organization in the health sector) must be willing to work with the nonprofit—and typically also an intermediary—to determine the specifics of a PFS deal, dedicate the time and energy to thinking about the value of key outcomes, and commit to negotiating these terms. At the same time, the intermediary must begin to engage with investors to understand whether the program can plausibly be supported by PFS capital.

# FEASIBILITY ASSESSMENT

Social Finance’s feasibility work with nonprofits, funders, and governments looks into these same areas, but in significantly more depth. Guided by local leaders and community priorities, we typically focus our efforts on a single program prioritized during landscaping work.

As we discussed in the previous section of this paper, we’re hopeful that the feasibility process can be helpful to nonprofit leaders as they review their own strategy, link their evidence to policy-relevant outcomes, and carefully define the value that their programs create for governments and communities.

**FIGURE 4** Feasibility study overview



## DEFINING OUTCOMES

According to the United Way, outcomes are “the benefits or changes for individuals or populations during or after participating in program activities.”

These may be changes in “knowledge, attitudes, skills, behaviors or condition. They are what participants know, think or can do; how they behave; or what their condition is, that is different following the program.”

In our feasibility work, once we have identified the target population, program, and service provider of interest, we begin the deeper due diligence process. That process begins, perhaps predictably, with evidence.

## *Evidence in Pay for Success feasibility studies*

In the social sector, ‘evidence’ is about the confidence we have in the causal link between a program and its outcomes.<sup>13</sup> Evidence is not defined by one golden academic report; rather, it’s a constellation of corroborating research and information about a program’s effectiveness. At Social Finance, we think of evidence through different lenses, balancing the information across each.

We look first to formal evaluations of a given program—studies measuring impact against carefully defined comparison groups, typically conducted by an independent third party. We complement those evaluations with others looking at similar programs to see if they, too, have demonstrated effects. Finally, we look at the

---

13 See sidebar; *A Guide to Developing an Outcome Logic Model and Measurement Plan*. United Way of Greater Richmond & Petersburg, 2015. Web. 2015 Oct. [http://www.yourunitedway.org/sites/uwaygrp.oneeach.org/files/Guide\\_for\\_Logic\\_Models\\_and\\_Measurements.pdf](http://www.yourunitedway.org/sites/uwaygrp.oneeach.org/files/Guide_for_Logic_Models_and_Measurements.pdf).



provider's operational data—the evidence an organization itself has collected about its participants.

In the past decade, we've seen a growing emphasis in the social sector on “evidence-based practices.” This is an important development. But it's also important to be careful with this language. The existence or quality of evidence is not a binary. Pay for Success requires a more nuanced view. We seek to know not just if a program works, but how well it works (its effect size), for which populations, in which settings, and on which specific outcomes over what period of time. Evidence quality exists along a spectrum.

### **FORMAL EVALUATION**

To construct a Pay for Success project, we need to understand the estimated effect size of a given program on one or more outcomes. We need to be able to model not just whether a program creates a positive impact on, for example, earnings, but how many incremental dollars per hour participants will earn, and how that difference will persist or change over time. Formal evaluations are incredibly helpful in this respect. They provide an estimate of the effect size and provide confidence intervals around that effect, while working to avoid potential biases.

### **SPEAKING OF BIAS**

Biases are systematic inaccuracies in the estimates of how programs influence outcomes. They can make some kinds of programs appear more effective than they really are, and others appear less effective. Take, for example, a workforce development program that places 80% of its participants into jobs after 6 months. Sounds pretty good. But in many ways, this information alone isn't material. The result is susceptible to confounding variables: maybe there was a major tightening of the labor market (an example of an exogenous factor that can influence a program's outcomes). Perhaps these participants began the program during an especially challenging life moment—say, moments after being laid off—but would probably improve even in the absence of the workforce program, versus another program working with the long-term unemployed. (This is an example of what we call mean reversion.) Underlying all of this is the simple fact that everyone who signed up for this workforce program signed up for this workforce program—and in doing so, exhibited an underlying motivation that might drive success (self-selection bias).

All of this is to say: looking at programs in isolation is usually not very telling. In job training programs, even control groups—those

for whom no intervention is applied—consistently experience significant increases in employment. Comparisons are important: an 80% job placement rate might be underwhelming for a group of Fortune 500 CEOs, and a 20% placement rate might be excellent for a group of long-term unemployed workers facing substance use challenges during a recession.

### **FINDING THE RIGHT COMPARISON**

Because of these biases, it's important to understand not just what happens to program participants, but also what would have happened had they not received an intervention. Of course, we can't both give, and not give, an intervention to the same people. The best we can do is to track people with similar backgrounds and motivations who do and do not participate in the program.

Randomized controlled trials (RCTs) are one time-tested way to do this. RCTs are generally regarded as the strongest method for evaluating a program's effectiveness.<sup>14,15</sup> They take groups of eligible participants and then split them up randomly, giving one group the intervention and not giving it to the other. Comparing the two groups allows researchers to control for selection biases—and to ultimately conclude that average differences in outcomes between control and experimental groups are attributable only to the program.<sup>16</sup>

However, RCTs can often be challenging to implement, time-consuming, and expensive. As a result, other kinds of evaluation designs are often proposed in their place. "Quasi-experimental" studies compare results from a given program against a carefully constructed comparison group, often using government-collected data to compare outcomes. That comparison group can vary from

---

14 *Which Study Designs Are Capable of Producing Valid Evidence About a Program's Effectiveness?* Coalition for Evidence-Based Policy, 2014 Oct. Web. 2015 Nov. <http://coalition4evidence.org/wp-content/uploads/2014/10/Which-Study-Designs-are-Capable-of-Producing-Valid-Evidence-of-Effectiveness.pdf>.

15 Evaluation method is not the only indicator of a strong study. We also look to other markers of evaluation quality: statistical power, sample size, control group definition and provisions, contamination, study attrition.

16 RCTs have many benefits, but their utility should not be overstated. Experiments test the impact of a specific program at a specific time and place among a specific population. As those things change, so too will the results of an evaluation. In a sector marked by shockingly few RCTs, practitioners must constantly make tradeoffs in understanding the value of, for example, an RCT conducted decades ago in a different state versus the value of alternative options for evaluating a program's effectiveness.



relatively general (e.g., comparing outcomes for a group of long-term unemployed to a statewide average) to more specific (e.g., a comparison group statistically chosen to match the experimental group in demographics, while also exhibiting similar motivations, such as a group of applicants accepted to a program who—for reasons beyond their control—were not able to attend).<sup>17</sup> The more specific the comparison, the more robust the findings.

Formal evaluations provide us the most predictive information about a program’s effectiveness. When it comes to evaluation, though, one size does not fit all; different kinds of analysis make sense for different programs.

In some cases, it may be reasonable to rely on evaluations without comparison groups. This is most likely to be true when positive outcomes are rare without intervention or when historical rates of outcomes are relatively stable.

#### **OTHER RELEVANT EVIDENCE**

We also look at the evaluations of other, similar programs. This is helpful in validating that a given program is likely to be effective. It can also provide valuable information on how to best collect information about impact: where to find the data, who the experts are in the field. Nevertheless, there are numerous examples of different “translations” of a given program—all working from a similar framework—returning very different results when applied in a different location or context.<sup>18</sup>

---

<sup>17</sup> Quasi-experimental designs can vary widely. Evaluations without comparison groups, such as studies comparing participants before and after an intervention (sometimes called “pre/post” designs), may be referred to as quasi-experimental. However, such studies are threatened by large external changes, self-selection bias, mean reversion, and maturation effects. More rigorous methodologies begin to account for some of these threats. Regressions and propensity score matching can statistically create comparison groups that share similar characteristics to the experimental group. However, these methods are still vulnerable to motivational biases. Other methods can define comparison groups that are nearly as good as random. For example, regression discontinuity designs, which look at the effects of an intervention on a population just above and below some kind of threshold or cutoff, may fall into this category. Nevertheless, no evaluation is perfect. All are vulnerable to errors in data measurement and the questions about their generalizability.

<sup>18</sup> See, for example, the widely varying impacts of community translations of the Diabetes Prevention Program.



We seek to know not just if a program works, but how well it works, for which populations, in which settings, and on which specific outcomes over what period of time.

In general, the more similar an evaluation is to the program in question, the more relevant it is. We typically think of relevance across three categories.

- Geography: evaluation taking place in the geography in question, or another that is very similar;
- Population: evaluation involves a treatment group similar to the organization's target population;
- Operator: evaluation conducted on the organization's own program, rather than on another party.

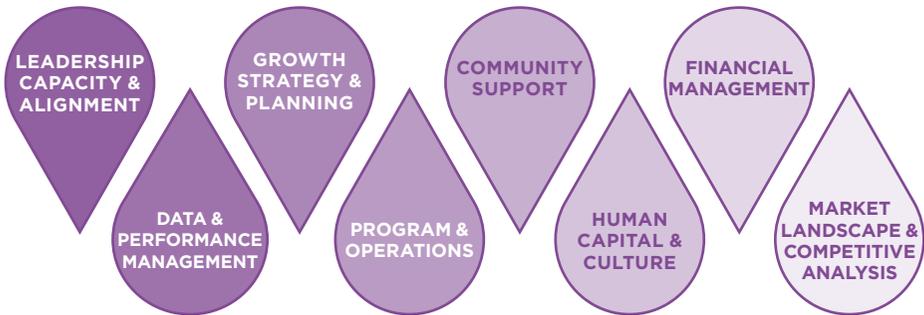
Evaluations that are more relevant—in that they meet the criteria outlined above—will improve the predictability of a program's impact.

### ***Organizational capacity assessment***

PFS projects typically require nonprofits to scale up their operations significantly within a short period of time. To successfully scale while maintaining quality requires strength across many dimensions.

With this in mind, Social Finance has developed a tool to help support the PFS due diligence process. We break these analyses into eight categories, seeking to understand the strengths and challenges an organization faces.

## ORGANIZATIONAL CAPACITY ASSESSMENTS: KEY ELEMENTS



**FIGURE 5** Advisory due diligence framework.

### LEADERSHIP CAPACITY AND ALIGNMENT

Great organizations have great management teams. We start by evaluating the executive team’s experience, their track record of managing growth and change, and their ability to think strategically and prioritize effectively. We similarly review the composition of the organization’s Board and the alignment between Board and senior management.

### GROWTH STRATEGY AND PLANNING

To scale via PFS, nonprofits need a clear vision and a significant opportunity for growth. We assess these qualities by looking at the size of the social issue the organization is seeking to address (the “market size”); their theory of change for how to impact that issue; a realistic growth strategy for how to expand; and the efficiencies the organization may be able to capture as it grows. At the same time, we look backwards to assess whether the organization has demonstrated its ability to scale—to what extent have services been expanded? With what results?

### COMMUNITY SUPPORT

While market size is important, we’ve found that there is often a major disconnect between the need for social services within a community and the demand for them. High demand for services,

and a proven ability to generate interest and recruit participants, is essential to successful scale-up. Interest is often driven by high levels of participant satisfaction, as well as strong connections with local leaders, relevant policy and advocacy efforts, and sector thought leaders. Well-managed external communications are also important.

### **FINANCIAL MANAGEMENT**

Financial stability, balance sheet strength, and strong external funding partnerships help provide confidence that the organization will be able to support a project through its conclusion. Consistent and clear budgeting, as well as robust processes and audits, establish the credibility of the organization's internal operations.

### **DATA AND PERFORMANCE MANAGEMENT**

Data, and the ability to use those data to troubleshoot and improve, are essential to top-quality nonprofits. In our diligence process, we look for robust and consistent data capture and cleaning, with key data aligned against the organization's theory of change. Quality control and fidelity monitoring ensure that programs will be delivered as anticipated. The capacity to deliver value-add data analyses demonstrates an organization's ability to proactively solve problems that may arise and course-correct to achieve its goals.

### **PROGRAM AND OPERATIONS**

To ensure that the program can consistently deliver its evaluated outcomes, it must be codified enough to be easily replicable, with strong internal processes and protocols clearly delineated and a target population that is well defined. Strong program managers carefully manage operating costs and continuously assess the key risks underlying their work.

### **HUMAN CAPITAL AND CULTURE**

Perhaps the most important indicator of organizational capacity, and one of the most challenging to measure, is the depth and breadth of organization talent. We look for strong processes (recruiting and retention plans, training, and professional development), clarity of roles and organizational structure, and a cohesive and supportive culture.

### **MARKET LANDSCAPE AND COMPETITIVE ANALYSIS**

Ultimately, the organization's value proposition must be compelling in comparison with others in the field. The organization should have an understanding of the competitive landscape and should be demonstrating success in expanding its share versus other, similar programs.

## *Understanding value*

At the core of a Pay for Success contract are the outcomes—how helping people can improve lives and, in the process of doing so, create social value. But selecting outcomes for a PFS project is not straightforward. The process is, by its nature, reductionist: it can feel like simplifying an entire program, and the program's wide-ranging benefits, to just two or three metrics. To make it even tougher, these outcomes are selected not only for the impact they have on people's lives, but also for the impact they have on a payor's budget or policy priorities—often within a short period of time.

So how do we describe the value of a particular outcome? We often think about this in a matrix. On one axis of the grid is time: whether the benefit of a program accrues within the lifespan of a given contract (“in-year” benefits) or whether it accrues thereafter (“out-year” benefits). Most payors are willing to pay for benefits that accrue within the project lifespan; they are often willing to pay for the present value of future benefits, but sometimes at a discounted rate.

On the other axis of the grid is the kind of value created: fiscal value, which may impact a payor's bottom line, versus community benefits, which more broadly impact the lives of those served (but not the payor's bottom line). Payors have widely varying perspectives on fiscal and community benefits. Most are willing to pay for the former. The latter—while representing the central purpose of government—are often the subject of greater negotiation: some governments are more or less willing to quantify and pay for the economic value of community benefits.

For this reason, we often begin conversations about economics with in-year fiscal benefits. In an education-focused PFS project, for example, these might include the savings a school district might realize from improving grade promotion in first and second grade following a successful pre-K program, or the value of special education utilization reductions. Looking over the longer term into out-year fiscal benefits, they might also include the present value of improved promotion in grades three through six, or lower disciplinary rates in middle school. We may also look at other community or social benefits of high-quality pre-K, such as better student achievement or graduations rates—outcomes most school districts care about deeply, but which might not directly impact their budgets. Ultimately, the total value of a set of outcomes in a PFS project might include elements of each of these. Outcome values may even extend beyond the value

FIGURE 6  
**ECONOMIC ANALYSIS FRAMEWORK**

EXAMPLE OUTCOMES FOR EDUCATIONAL PROGRAM



## TRACING FUNDING FLOWS TO PAYORS

Unlike government agencies, people’s lives defy neat categorization. An early childhood development program may reduce emergency room visits, while at the same time limiting reliance on special education programs. Should the department of health act as payor, or the department of children and families, or the department of education? These kinds of challenges, spread across agencies in the same jurisdiction, are often called “horizontal” problems. But even some programs that fall entirely within the purview of a single department may be complicated because the relevant funding comes from a blend of local, state, and federal payors, creating so-called “vertical” problems. Children less likely to require special education are less expensive to educate—if education funding for that child comes not only from the town, but also from the state and the federal government, who is the right payor?

Funding flows between different levels of government (federal, state, and local) or between different government agencies (such as a state department of corrections and a state department of health) can complicate the picture. In our feasibility work, we are careful to understand not just how much benefit is created, but where that benefit accrues, in order to identify where the greatest amount of value accrues.

accrued to an individual payor—as when a pre-K program improves non-scholastic outcomes, such as health, productivity, or public safety.

It's important to realize that a lot of assumptions go into any estimation of value—ranging from which outcomes to include, to the program's effect on those outcomes, to the relative weighting of short- and longer-term outcomes. Reasonable estimates can vary; the key point is that this kind of analysis can help to make the case for intervention more tangible, highlighting when the benefits of a program are greater than the costs of providing it.

## *Identifying payors*

Nonprofits create value for a range of stakeholders. Once we've identified the value a program creates, we attempt to understand for whom this value is most relevant. To do so, we map a program's outcomes to local public policy interests or existing initiatives across potential payors—local, state, or federal governments, employers, healthcare organizations, or others who might benefit from positive outcomes. We quantify where different fiscal and community benefits accrue within those payors' budgets and priorities, and whether they accrue to a single agency, or to multiple.

Often in this process we begin to engage with payors to understand their interest and willingness to pay for different outcomes. Building on our work in understanding the cost of the proposed program, we help payors understand how a PFS project may influence their budgets and their broader policy goals.



# IMPROVING PERFORMANCE TO IMPROVE LIVES

---

**WE WANT OUTCOMES—BETTER HEALTH, STRONGER FAMILIES,** a more just and peaceful society—but too often we pay for inputs: for numbers of visits, for pamphlets distributed, for therapy provided.

There are more thoughtful ways to fund services. And there is a growing coalition behind the incredibly simple idea that we should spend taxpayer and philanthropic money on programs that work—particularly when it comes to services for the most vulnerable.

A central principle of that movement, we believe, should be to fund programs based on their performance. Results matter. By paying for them, we can encourage and expand programs that work and move closer to breaking cycles of intergenerational poverty.

Performance-based funding in its own right is important. But directing that funding to organizations with the best evidence and ability to grow—to what works—is better still. Defining what works is complex. It includes reviewing formal research, but also understanding organizations' own experiences, their capacity to deliver services well, and their ability to expand. Guiding funding decisions with thorough due diligence, too, helps to strengthen the sector and improve the lives of beneficiaries.

Ultimately, we believe that Pay for Success can strengthen these parallel movements. It creates performance-based funding structures and guides policymakers and philanthropists toward organizations that make the most impact with those tools. At the same time, it helps nonprofits to think critically about the evolution and development of their organizations.

We believe that the social sector is moving—slowly, but inexorably—toward a greater alignment between performance and payment. The consequences of this movement are momentous. We are hopeful that, as the sector changes, great organizations will finally get the support they need to reach scale—and we're hopeful, too, that the tools of Pay for Success can accelerate them along that path.

