

Learning Social Metrics from International Development

Paul Veldman

Columbia University

The goal of community development should be to improve the lives of as many disadvantaged people as possible with a minimum amount of resources. The field can achieve this goal by identifying those community development projects that are both efficient and have significant social impacts, and then scaling up. However, without effective evaluations, it is impossible to gauge progress. Effective evaluations include randomized experiments and well-designed observational studies, both of which can measure the impact to society. This article analyzes the best evaluation methods in international development, including examples of evaluation toolkits, with the aim of helping community development practitioners apply the methods to projects in the United States. The article also includes steps and recommendations to increase the quantity and quality of evaluations for community development projects.

Introduction

In many respects, the U.S. community development field lags the international development community in the number and precision of evaluations methods used to measure social impacts. Evaluations are costly in both time and resources, and they can divert social enterprises from delivering valuable services or goods to clients. However, sound evaluations result in information that can help guide more effective development projects while using funds more efficiently. Furthermore, when the results are publicly shared, the entire development community benefits. In the long run, quality evaluations will result in more development goals being achieved for less money.

With high “opportunity costs,” evaluation research becomes a balance between collecting valuable information to improve operations and delivering goods and services to a target population. In contrast, in the for-profit arena, the market often automatically generates the information needed for an industry to evaluate its operations and make the decisions that allow it to operate more efficiently. Although some market indicators can be used to improve the efficiency of development projects, these projects are usually implemented precisely to address the breakdown of information that results from market failures.

In the United States, several programs and policies support antipoverty and community development programs, including the Community Reinvestment Act of 1977 (CRA). CRA redresses the market failure of inadequate access to capital in low-income areas. The CRA requires regulated financial institutions in the United States to meet the credit needs of

the local communities where they do business. The CRA investments are often funded in conjunction with nonprofits and federal, state, and local governments.

CRA investments have allocated billions of dollars to community development projects across the United States. However, the social impact of these investments is rarely evaluated. Those evaluations that have been done are generally so poorly executed that it is difficult to assess whether the goals of the program are being addressed, let alone achieved. Declaring that an investment has allowed a school to build new classrooms reveals nothing about whether school attendance has increased, test scores have increased, or graduation rates have improved. Even if the goals are met, a proper evaluation is still needed to determine whether other, less-expensive and more efficient interventions would have achieved the same outcomes. Rigorous evaluation also reveals the reason why a certain project or policy succeeded or failed, which is valuable information for future development practitioners.

To date, evaluations in U.S. development projects have been underfunded, for a variety of reasons. Many government departments and institutions, for example, do not require them. Evaluations are technically complex and can raise questions of data accuracy and bias in reporting. Further, it is difficult to standardize evaluations, which reduces the ability of third parties to understand and use the data. Evaluations can also be politically damaging, particularly if the results are negative. Methodologically, quality data are often unavailable;¹ it is sometimes difficult to account for externalities; and selection biases in funding can confound results. Finally, there is a general lack of consideration for long-term analysis, and there are few incentives to develop new tools needed for evaluations.²

A further problem with evaluations is that practitioners want rapid feedback, both to make managerial decisions and to appease donors. However, researchers often need much more time to collect and analyze data. It can sometimes take years before certain social impacts from a program can be measured. This “myopia bias”—the tendency for short-term impacts to be more readily studied than long-term impacts—inevitably leads to the funding of development projects that provide short-term results, which are often less effective.³

In addition, the market does not accurately value the public benefits—positive externalities—of evaluations. Practitioners would benefit from more data because they could design more effective development projects when using the results from the evaluation of other projects. This added value is often not considered sufficiently important to incur the costs for the evaluation. Also, evaluation research is underfunded because development impacts are difficult to identify and measure. Most development projects have spillover effects that

1 Judy L. Baker, *Evaluating the Impact of Development Projects on Poverty: Handbook for Practitioners* (Washington, DC: LCSPR/PRMPO, The World Bank, 2000). See also Lisa A. Hagerman and Janneke Ratcliffe, “Increasing Access to Capital: Could Better Measurement of Social and Environmental Outcomes Entice More Institutional Investment Capital into Underserved Communities,” Federal Reserve Bank of San Francisco, *Community Development Investment Review*, vol. 6, Issue 1, (2010).

2 Martin Ravallion, “Evaluation in the Practice of Development,” (New York: World Bank, March 2008). *The World Bank Research Observer*, vol. 24, no. 1 (February 2009)

3 Ibid.

affect the surrounding community. These effects are often unintended and not measured. More focused development projects, with well-defined beneficiaries, are more likely to be thoroughly evaluated. The evaluation field is further complicated by the tendency of development practitioners to disguise advocacy information as evaluation research. They often only highlight successful projects and ignore failures for the purpose of attracting donors.⁴

However, attitudes are changing. Nonprofits, government agencies, foundations, and social entrepreneurs are realizing the benefits of evaluation and the need to address obstacles to evaluation. The various groups have developed several cost-effective evaluation methods and toolkits, some more rigorous than others. This paper discusses five methods of effective evaluation: randomized evaluation, observational evaluation, cost-benefit analysis, citizen participation, and process evaluation. Before discussing these methods in more detail, I discuss several aspects that all well designed evaluations share.

Best Practices of Evaluation Methods

Evaluation research provides greater value when it demonstrates causality. Causality is the relationship of an intervention to a change in a society, where the change in society is a consequence of the intervention. In other words, the intervention is directly responsible for the change. It is isolating this causality that is difficult, if not impossible, and usually expensive. Isolating causality requires controlling for a large number of variables, which is difficult when analyzing a community or society.

A development project creates a chain of events, referred to as the “results hierarchy.”⁵ Each part of the chain can be evaluated to provide information on the effects of the project. The sequence of the hierarchy is:⁶

1. Inputs—the resources employed to implement activities.
2. Activities—the actions taken by an organization.
3. Outputs—the direct and early results of the activities.
4. Outcomes—the intermediate observable and measureable changes that serve as progress toward an impact for the target population.
5. Impacts—the ultimate sustainable changes sometimes attributed to an action or combination of actions.

Evaluation research is designed to measure these various factors. The challenge is that researchers must collect enough data and control for enough variables to demonstrate the causality of a project. Furthermore, the value of the research is enhanced if the findings can be applied to other projects and policies. However, as the focus of a development project

4 Ibid.

5 Fay Twersky, Amy Ratcliffe, Jodi Nelson, Phillip Setel, and Lance Potter, “A Guide to Actionable Measurement” (Seattle: Bill and Melinda Gates Foundation, 2010).

6 Ibid.

narrows in order to control for variables and to generate more accurate evaluation research, the less applicable the project intervention is for development projects in other communities.

The five methodologies we review here share five components that are necessary to address in a thorough evaluation: time horizon, sample size, transparency, third-party evaluation, and counterfactuals. These are of special importance because they are the components that are most often ignored or not properly incorporated into evaluations.

Time Horizon. An evaluation should be adjusted for the appropriate timeline of a project. Ideally, the evaluation methodology also should include periodic monitoring that will provide updated information about the social impacts of a project. Designing the timeframe poses problems as some projects can take years or even decades for the social impact to reach the point where it can be measured. For example, the evaluation of a project that improves a community's drinking water will have a much shorter timeframe than an early childhood development program.

Sample Size. The Social Development Family of the World Bank analyzed 60 evaluations of community-driven development (CDD) projects. The majority of the evaluations, the assessment found, suffered from limited scale, meaning that the evaluations lacked the sample size necessary to generate statistically significant results.⁷ A mathematical formula that uses population size and variance should be used to determine a sample size that is sufficiently large for the analyses to be sensitive enough to detect statistically significant results.

Transparency. Transparency includes not only disclosing the results of the evaluation but also clearly describing the methodology used to collect and analyze the data. This includes explaining the rationale behind all assumptions made, equations used, social indicators measured, sample population selected, time period analyzed, counterfactuals used, and so forth. Transparency is a key component to any thorough evaluation. First, it allows other development practitioners to identify which interventions did and did not work, and often why. Second, it allows the data and analysis to be verified. Third, it allows evaluations to be accurately replicated when applied to other similar projects. Fourth, transparency provides the opportunity for practitioners to identify flaws in an evaluation and to improve later implementations.

Third-Party Evaluation. A third-party evaluation increases validity. It helps ensure that the data are accurately collected and not influenced by any conflicting interest of the organization being evaluated. However, third-party evaluations are usually more expensive. A less costly option is to have a third party audit a self-conducted evaluation of an organization.

Counterfactuals. To isolate the impact of a development project, it is necessary to estimate the counterfactual; that is, what change, if any, would have occurred if the project were never implemented. For example, presume the income of an individual participating in a

7 Paul Wassenich and Katherine Whiteside, "CDD Impact Assessments Study: Optimizing Evaluation Design Under Constraints" (New York: Social Development Family, World Bank, February 2004).

microfinancing program increases. Is this increase because of the microfinancing program or factors in the community, such as a newly constructed road linking the community to new markets? To identify the counterfactual, it is necessary to “net out” the effect of the intervention from other variables. This is accomplished by comparing control groups (those that do not participate in the intervention) with treatment groups (those that do participate in the intervention).⁸ Although counterfactuals are necessary to estimate the true impact of a development project, they are often ignored. Setting up, monitoring, and evaluating counterfactuals requires time and resources that could otherwise be used for delivering services or goods.⁹ In addition, donors and investors typically prefer that their funds be allocated to operational activities and not to analyzing counterfactuals that do not directly benefit the clients or beneficiaries. Additionally, it could be unethical to withhold a program to the counterfactual if resources are available.

The world of international development evaluations is a complex one, and as with many complex topics, an extensive spectrum of tools has been developed to shed light on the subject. This is not surprising when considering that evaluations must control for a seemingly infinite number of variables, in thousand of locations around the world, and do so with varying budgets. I have categorized these evaluation tools into the five types of evaluations noted above with the aim of helping the field better understand how they work, what they can tell us, when they should be used, and how difficult and costly they are to implement.

I begin with the most rigorous technique, randomized evaluations. Starting with the “gold standard” will be helpful in understanding the other categories of evaluation if this method proves too costly or logistically infeasible. Randomized evaluations are important because they isolate causality.

Overview of Evaluation Tools Used Internationally

Randomized Evaluations

In this method, researchers randomly assign participants (the sample) to a treatment group or a control. The treatment group receives the intervention and the average impacts are compared with those in the control group. This method bypasses several weaknesses of other methods. If the sample size is large enough, the attributes of the sample will represent the attributes of the entire population. In other words, what is true for the sample size is most likely true for the entire population. In addition, when individuals are randomly assigned to two groups of sufficient size, they are considered statistically identical to the larger population and to each other. This provides an ideal counterfactual, as one group known (treatment group) receives development interventions and the other group (control group) does not. All other variables are controlled for, meaning they will have equal influence on the outcomes

⁸ Baker, *Evaluating the Impact of Development Projects on Poverty*.

⁹ Claire Brown, “Developing an Effective-Efficient Assessment Template for Social Enterprises” (forthcoming, early manuscript provided to the author, June 2010).

and thereby the effects of the intervention are isolated. The impact is then determined by measuring the differences in outcomes between the control group and the treatment group for a period of time after an intervention.

In the international development field academic organizations such as Massachusetts Institute of Technology's Abdul Latif Jameel Poverty Action Lab (J-PAL), and University of California, Berkeley's, Center for the Evaluation of Global Action (CEGA) lead the field in using randomized testing. In addition, other organizations, such as the Bill and Melinda Gates Foundation and the William and Flora Hewlett Foundation, are realizing the importance of randomized evaluations and are allocating more resources to such evaluations.

A widely cited success in randomized evaluation is CEGA's Primary School Deworming Project (PSDP). The PSDP evaluated school-based mass treatment with deworming drugs in Kenya. The evaluation phased in the program in three stages at three sets of randomly selected primary schools. The evaluation found that the program reduced absenteeism and increased school performance in treatment schools by 25 percent, and the program was much cheaper than alternative methods for boosting school participation, such as purchasing new textbooks.¹⁰ The program also had sizable "spillover" effects. The health and school participation improved for children who were not taking part in the experiment in both the treatment schools and neighboring schools owing to fewer instances of worms being transferred among students.¹¹ In this example, not only can randomization accurately show program impacts, but also it can reveal seemingly unlikely solutions to difficult problems. This opens doors for creative development solutions to be tested.

Randomization has its limits, however. First, this method can be costly. It can also require modification to the implementation plan, and sometimes these adjustments are deemed unacceptable. Third, some programs are so idiosyncratic that conducting a randomized evaluation would not generate useful, widely applicable information. Fourth, some programs and sectors do not have a large enough control population or a large enough sample size.¹² Finally, it might be unethical or politically infeasible to deny a program to a control group.¹³

Randomized evaluations must be specifically designed for each development project, and therefore there are no existing general toolkits that can be applied to every randomized evaluation. However, J-PAL provides an abundance of material and information describing randomized evaluation methodology and best practices, including information on planning, implementing, obtaining results, and how to draw policy conclusions from randomized evaluations.¹⁴

10 David Leonhardt, "Making Economics Relevant Again," *New York Times*, February 20, 2008.

11 Edward Miguel and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72 (1) (January, 2004).

12 Franck S. Wiebe, Millennium Challenge Corporation, "Working Paper: Aid Effectiveness: Putting Results at the Forefront" (Washington, DC: Millennium Challenge Corporation, October 27, 2008).

13 Abdul Latif Jameel Poverty Action Lab, "When is Randomizing (Not) Appropriate?" (Cambridge, MA: Massachusetts Institute of Technology, 2010), available at www.povertyactionlab.org/methodology/when/when-randomization-not-appropriate.

14 Abdul Latif Jameel Poverty Action Lab, "Methodology: Overview" (Cambridge, MA: Massachusetts Institute of Technology, n.d.), available at www.povertyactionlab.org/methodology.

Randomization is the most accurate way to create a counterfactual but other methods are available. These methods of creating a comparison group can provide useful insight, but they rely more on assumptions. It is usually impossible, and always difficult, to ensure that all the assumptions are true, and it is these assumptions that usually cause debate about a project's claimed social impact.¹⁵

Observational Evaluations

If proving causality is not feasible, the next best thing is to illustrate correlation between a particular action and a set of outcomes.¹⁶ This is what observational evaluations can provide. Observational evaluations can illustrate, for example, that distributing mosquito nets is highly correlated with fewer malaria cases. It is still not known how much the mosquito nets contributed (or if it even did) to the reduction of malaria given that other variables could simultaneously be influencing the malaria rate. However, the higher the correlation, the more likely it is that the mosquito nets are having an impact. What must then be decided is if the correlation is high enough to justify spending resources to distribute mosquito nets. This can only be determined by conducting a cost-benefit analysis (see below).

There are three main types of observational evaluations, also referred to as quasi-experimental evaluations. The simplest compares the same group before and after an intervention, also referred to as pre-post evaluation. This method requires the assumption that the intervention was the only factor (or at least the major factor) influencing any changes in the measured outcomes over time.

A second method uses a cross-sectional approach and compares the treatment group with a similar group that exists elsewhere and that is not created through randomization.¹⁷ This approach can be further divided into what is called either simple difference or differences in differences. Simple difference is the measure of difference after the intervention is implemented between those who receive an intervention and those who do not. This requires the assumption that nonparticipants are identical to participants except for the intervention, and that both groups were equally likely to receive the intervention before it was implemented. Differences in differences measures the improvement over time of those who received an intervention relative to those who did not. The assumption of this method is that if the program did not exist, the two groups would have identical trajectories over the specified period.¹⁸

As mentioned above, these nonrandomized evaluations cannot provide proof of causality, but they can provide evidence of correlation between an intervention and a change in a social outcome. To revisit the example used above, supplying mosquito nets to a community could be correlated with a decline in malaria cases, but this could be attributed to other variables

15 Abdul Latif Jameel Poverty Action Lab, "When is Randomizing (Not) Appropriate?"

16 Examples of observational evaluations can be found in Appendix A.

17 Paul Brest and Hal Harvey, "Money Well Spent: A Strategic Plan for Smart Philanthropy" (New York: Bloomberg Press, 2008).

18 Abdul Latif Jameel Poverty Action Lab, "When is Randomizing (Not) Appropriate?"

that are not accounted for in the evaluation, such as a decline in the mosquito population. However, observational evaluations do provide useful information and can illustrate if an organization is heading in the right direction with its interventions. Correlation combined with one's knowledge of how the world works, perhaps gained from more thorough evaluations of similar interventions, can at least hint strongly at causation.¹⁹

Observational evaluation methods rely heavily on the collection of baseline data. Baseline data allow for the comparison of social indicators before and after a development project is implemented. This provides additional information about how a project has affected a target population. However, comparing baseline data with results after the project's implementation does not necessarily provide an accurate measure of the project effect owing to confounding variables that can influence the target population. A counterfactual would be needed to eliminate or reduce the possibility that other variables may be affecting a target population. Nevertheless, baseline data are valuable in comparisons, particularly between various development projects that target populations with different characteristics or that live in other geographic locations.

Baseline data provide information on the level of education, infrastructure, income, and other factors, all of which are necessary for designing an effective development project. A given development project might have been successful with one target population, but the results could be very different in a community with other characteristics. The more baseline data collected, the better the understanding of how various development projects affect different populations. Again, this data collection takes time and money, and the feasibility of collecting the data must be considered in the design of the evaluation.

There is another category of observational evaluations that includes a subset of models that rely on regression analysis to determine correlation between interventions and outcomes. Regression analysis includes techniques for modeling and analyzing the correlation of one or more variables (independent variables) with a specific outcome (dependent variable). This type of analysis provides insight on how the value of the dependent variable changes on average when any of the independent variables are varied. The independent variables can be further categorized as control variables and noncontrol variables. The control variables are included to isolate the noncontrol variables' correlation with the dependent variable and to exclude their impacts from the measured changes in the dependent variable. The inclusion of control variables assists in isolating the correlation and influence of an independent variable of interest relative to the dependent variable.

For example, with a sufficiently large sample size, a regression equation could be calculated where the dependent variable is the number of malaria cases and the independent variables are the number of mosquito nets distributed, average rain fall for the year, community population, and average level of education. The average rainfall is used to control for the number of mosquitoes that exist which is largely based on amount of rainfall, population size controls for the ease of malaria to be transmitted from one person to another, and the

19 Brest and Harvey, "Money Well Spent."

level of education is used as a proxy for the level of understanding of how malaria is transmitted. Creating such a regression equation allows an evaluator to say something similar to “on average, a distribution of 100 additional mosquito nets is associated with a decrease of 50 malaria cases a year, holding the average rainfall, population size, and level of education constant.” The larger the sample size and the more relevant the controlling variables added to the equations are, the greater the accuracy of the correlation will be between the dependent variable (number of malaria cases) and the independent variable of interest (number of mosquito nets distributed). However, as with all observational evaluations, regression analysis does not prove causality, but it can illustrate high levels of correlation that suggest causality.

Regression analysis can be extremely technical and be used in a variety of ways. Some of the subcategories of regression analysis used to analyze development work include multivariate regression, regression discontinuity design, statistical matching, and instrumental variables (see Table 1).

Multivariate regression compares individuals who receive treatment against those who do not, and other factors that might explain differences in the outcomes are accounted for with control variables. This method requires that information for both participants and nonparticipants be collected before and after the intervention.

Regression discontinuity design ranks individuals on the basis of specific measurable criteria. A cut-off is created that determines whether an individual is eligible to participate in a program or not. The comparison group is composed of individuals who are close to the cut-off point, but who fall on the “wrong” side of that cut-off and therefore do not receive the program. The experiment group consists of those who were close to the cut-off point, but who fall on the “right” side. This allows for criteria variables to be very similar between the participants and nonparticipants, and therefore these variables are somewhat controlled.

Statistical matching uses regression analysis to compare individuals in the treatment group with similar individuals in the control group. This is achieved by either matching participants from each group who have exactly the same selected characteristics, or who have a mix of characteristics that predict that they would be as likely to participate in the program.

Instrumental variables analysis uses an almost random factor (instrumental variable) that is uncorrelated with the outcomes other than the fact that it predicts an individual’s participation. The control group is composed of individuals who, because of this close to random factor, are predicted not to participate (and possibly as a result) do not participate. For example a researcher may attempt to estimate the effect of smoking on health from observational data by using the tax rate on tobacco products as an instrument variable for smoking in a health regression. If a tobacco tax only affects health because it affects smoking (holding the other variables in the model constant), correlation between a tobacco tax and health is evidence that smoking causes changes in health.²⁰

20 Jonathan Gruber and Botond Koszegi, “Tax Incidence When Individuals Are Time-Inconsistent: The Case of Cigarette Excise Taxes” *Journal of Public Economics*, 88 (nos. 9-10) (August 2004): 1959-1987.

Table 1. *Experimental Evaluation Methods*

Methodology	Description	Who is in the Comparison Group?	Required Assumptions	Required Data
Pre-Post	Measure how program participants improved (or changed) over time.	Program participants themselves—before participating in the program	The program was the only factor influencing any changes over time.	Before and after data for program participants.
Simple Difference	Measure difference between program participants and non-participants after the program is completed.	Individuals who didn't participate in the program (for any reason), but for whom data were collected after the program.	Non-participants are identical to participants except for program participation, and were equally likely to enter program before it started.	After data for program participants and non-participants.
Differences in Differences	Measure improvement (change) over time of program participants relative to the improvement (change) of non-participants.	Individuals who didn't participate in the program (for any reason), but for whom data were collected both before and after the program.	If the program didn't exist, the two groups would have had identical trajectories over this period.	Before and after data for both participants and non-participants.
Multivariate Regression	Individuals who received treatment are compared with those who did not, and other factors that might explain differences in the outcomes are “controlled” for.	Individuals who didn't participate in the program (for any reason), but for whom data were collected both before and after the program. In this case data is not comprised of just indicators of outcomes, but other “explanatory” variables as well.	The factors that were <i>excluded</i> (because they are unobservable and/or have not been measured) do not bias results because they are either uncorrelated with the outcome or do not differ between participants and non-participants.	Outcomes as well as “control variables” for both participants and non-participants.
Statistical Matching	Individuals in control group are compared to similar individuals in experimental group.	Exact matching: For each participant, at least one non-participant who is identical on <i>selected characteristics</i> . <u>Propensity score matching</u> : non-participants who have a mix of characteristics which predict that they would be as likely to participate as participants.	The factors that were <i>excluded</i> (because they are unobservable and/or have not been measured) do not bias results because they are either uncorrelated with the outcome <u>or</u> do not differ between participants and non-participants.	Outcomes as well as “variables for matching” for both participants and non-participants.

Quasi-Experimental Methods

	Regression Discontinuity Design	Individuals who are close to the cutoff, but fall on the “wrong” side of that cutoff, and therefore do not get the program.	After controlling for the criteria (and other measures of choice), the remaining differences between individuals directly below and directly above the cutoff score are not statistically significant and will not bias the results. A necessary but sufficient requirement for this is that the cutoff criteria are strictly adhered to.	Outcomes as well as measures on criteria (and any other controls).
	Instrumental Variables	Individuals who, because of this close to random factor, are predicted not to participate and (possible as a result) did not participate.	If it weren’t for the instrumental variable’s ability to predict participation, this “instrument” would otherwise have no effect on, or be uncorrelated with, the outcome.	Outcomes, the “instrument,” and other control variables
Experimental Method	Randomized Evaluation	Participants are randomly assigned to the control groups.	Randomization “worked.” That is, the two groups are statistically identical (on observed and unobserved factors).	Outcome data for control and experimental groups. Control variables can help absorb variance and improve “power.”

Source: Abdul Latif Jameel Poverty Action Lab, available at [www.povertyactionlab.org/sites/default/files/documents/Experimental percent20Methodology percent20Table.pdf](http://www.povertyactionlab.org/sites/default/files/documents/Experimental%20Methodology%20Table.pdf).

Appendix A describes a number of internationally used toolkits that are used to measure observational data from development projects. These examples provide creative and useful techniques that could be adapted to measuring domestic community development projects.

Cost-Benefit Analysis

Although cost-benefit analysis is not mutually exclusive from randomized and observational evaluations, it stands as its own category as well.²¹ Both randomized and observational evaluations should include some sort of a cost-benefit analysis because it answers the question “what development impact does \$1 of spending achieve?” It is only when this question is answered that the efficiency of a development program can be understood, and can, consequently be gauged against other development projects.

It is very difficult to quantify social impacts in monetary terms, mainly because these determinations are often subjective. This subjectivity can result in debates over the accuracy of measurements. Cost-effectiveness analysis can make this process clearer. This method divides the impact of a program (e.g., the increase in the number of students attending school) by the cost of the project. However, it is still difficult to compare different projects that are addressing various aspects of development.²² For example, which is more valuable to society, 50 job placements or one more high school graduate?²³

In a study prepared for the Bill and Melinda Gates Foundation, analysts compared eight integrated cost approaches to measuring or estimating social value. In reviewing these different methods Paul Brest, President of the William and Flora Hewlett Foundation, stated that, in essence, all the methods are about expected return:

$$\text{Expected Return} = \frac{(\text{Outcome or Benefit} \times \text{Probability of Success})}{\text{Cost}}$$

As the study notes, the above simplification is useful, but each of the methodologies (REDF SROI, Robin Hood Foundation Benefit-Cost Ratio, Acumen Fund BANCO Ratio, Hewlett Foundation Expected Return, Center for High Impact Philanthropy Cost per Impact, and Foundation Investment Bubble Chart) answers the following questions differently:²⁴

- How are the outcomes or benefits estimated?
- How are the costs calculated?

21 Examples of cost-benefit analysis can be found in Appendix B.

22 Abdul Latif Jameel Poverty Action Lab, “When Cost-Benefit/Effectiveness/Comparison Analysis” (Cambridge, MA: Massachusetts Institute of Technology, n.d.), available at www.povertyactionlab.org/methodology/what-evaluation/cost-benefit/effectiveness/comparison-analyses.

23 Michael M. Weinstein, “Measuring Success: How Robin Hood Estimates the Impact of Grants” (New York: Robin Hood Foundation, November 12, 2007).

24 Melinda T. Tuan, “Measuring and/or Estimating Social Value Creation: Insights Into Eight Integrated Cost Approaches” (Seattle: Bill and Melinda Gates Foundation, December 15, 2008).

- How are uncertainties and partial attribution of result accounted for?
- How are the outcomes or benefits translated into natural units or monetized?

These cost-benefit evaluations attempt to create a counterfactual by estimating what would have happened to the beneficiaries of a program if they did not receive the program. For example, in estimating the cost-benefit of a job training program, the cost of the program is the dominator, and the number of beneficiaries who achieved job placement (the numerator) are multiplied by the their expected lifetime income, minus the average income they would have received if they had not received the job training. However, unlike randomized evaluations, this method does not consider the possibility that some beneficiaries could have become employed without the training. This issue can be somewhat addressed by subtracting an estimated percentage of the beneficiaries who would likely found a job without the training. However, as with many aspects of cost-benefit analysis for social impacts, this introduces assumptions that are vulnerable to credibility questions.

These cost-benefit evaluations also must make assumptions in order to estimate the present value of future benefits. The social impact of a program on a beneficiary should be estimated over his or her lifetime. However, estimating this amount can be difficult because the dollar value of a benefit received in the future is less than the dollar value of a benefit received today. This is because people in general prefer to receive something now as opposed to in the future and therefore, future costs and benefits must be discounted to present value. In financial terms this is referred to as the “discount rate.” However, there is no agreed upon method of estimating the discount rate for the many different social impacts.²⁵

Another assumption is applying a dollar value to the outcomes of social programs. This is referred to as the “shadow price.” Sometimes the shadow price is relatively easy to estimate; for example, the expected lifetime increase of attaining a job can be calculated by multiplying the average salary increase from the job placement by the average number of years that individual will most likely work. However, some impacts are more difficult to estimate. For example, there is no market value for the cost of crime to a victim. Similarly, value judgments must often be made. For example, is it better to save or improve the life of a 30-year-old as opposed to a 70-year-old?²⁶

Yet another aspect of social impacts to consider is that of interdependencies—when the outcomes of one or a series of interventions are dependent on other interventions. Interdependencies have an important influence on social impacts, yet they are very difficult to measure. For example, the increase in income levels from a microfinance program is also dependent on the level of infrastructure improvements, such as new roads. The success of an intervention will depend on many factors that are not addressed by the intervention. The number and influence of these interdependencies increases as the timeframe of the projected benefits increases.²⁷

25 Weinstein, “Measuring Success.”

26 Tuan, “Measuring and/or Estimating Social Value Creation.”

27 Ibid.

Even with all the above-mentioned assumptions and susceptibility for debate, cost-benefit analysis is useful and important tool for social impact evaluations. As Paul Brest, *Measuring and/or Estimating Social Value Creation: Insight Into Eight Integrated Cost Approaches*, explains, “You’re putting in a lot of incredibly speculative numbers, but ‘doing the numbers’ presses program officers to test their intuitions, and that’s likely to sharpen them.”²⁸ For cost-benefit analysis to be understood and used, it is important that the assumptions are clearly explained.

Appendix B provides six summaries of cost-benefit analysis methods used by leading non-profits and foundations. Each of the examples has their own method to address the difficulty of measuring and monetizing the social impacts from development interventions.

Citizen Participation

Yet another category of international development evaluation tools is evaluations that focus on citizen participation. These methods rely on the citizens, who ultimately are the beneficiaries of the development projects, to act as the evaluators. Asking citizens how development projects have affected them or which aspect of their lives they would like to improve can reduce the time and resources spent trying to answer these particular questions.

Several methods provide citizens with a greater opportunity to express their opinions on how public funds should be allocated. Two methods that have received international attention are participatory budgeting and citizen report cards.

Participatory budgeting. Participatory budgeting (PB) is a democratic process in which citizens meet to decide how to allocate a portion of a municipal or public budget and to identify, discuss, and prioritize public spending projects. PB attempts to improve both the performance and accountability of a bureaucracy that is not effectively representing the people and is underperforming. PB creates opportunities for engaging, educating, and empowering citizens. Also, PB addresses corruption by including citizens in the budget process, which promotes transparency and increases the accountability of government agencies. PB programs have typically involved individuals with low incomes and low levels of formal education, and as a result PB offers individuals from historically marginalized and excluded groups the opportunity to determine how resources are allocated. This typically results in funding being allocated to infrastructure and services that are more relevant to the communities they serve.²⁹

The first full PB process was developed in Porto Alegre, Brazil, in 1989. The World Bank’s evaluation of the program concluded that the PB led to direct improvements of facilities such as sewer and water connections. The program also resulted in an increase in the number of citizens who were participating in the budgeting process over the 10-year period of the study. In addition, the percentage of the budget that was allocated to both

28 Paul Brest to Melinda Tuan, “Re: Hewlett Foundation’s Return Methodology” (April 24, 2008) as cited in Tuan, “Measuring and/or Estimating Social Value Creation.”

29 Anwar Shah, ed., *Public Governance and Accountability Series: Participatory Budgeting* (New York: World Bank, June 9, 2010).

health and education increased substantially, from single digits to greater than 70 percent. These successes led to more than 140 (about 2.5 percent) of the 5,571 municipalities in Brazil adopting PB programs.³⁰

PB is now practiced in countries around the world, including Canada, which has implemented PB for public housing, neighborhood groups, and public schools, and in the United States in Chicago.³¹

Citizen report cards. Citizen Report Cards (CRC), like PB, provide an opportunity for marginalized and excluded citizens to voice their opinion about their underserved and neglected communities. The CRC process begins with a survey sent to a random set of citizens in a particular community, asking them to rate various public services and facilities. CRC not only provides accountability for government spending, which leads to more transparency and less corruption, but it also strengthens civil society through increased citizen influence over political decisions. CRC were first implemented in Bangalore, India, in 1993. Analysis of the subsequent CRCs reveals a decrease in corruption and a substantial increase in citizen satisfaction for most government agencies.³²

Process Evaluations

A final method for performing evaluations focuses on measuring an organization's internal management systems, processes and procedures, and incentives.³³ This methodology focuses on evaluating inputs and processes as opposed to outcomes. The logic is that if stakeholders can identify key elements or processes within organizations that result in successful development projects, then those key elements can be evaluated instead of project outcomes.³⁴ Measuring such indicators is usually less costly and time-consuming than trying to measure a variety of outcomes. This process has led to various certification and scoring systems for organizations.

One organization that has developed such an evaluation process is CGAP, with its SmartAid methodology.³⁵ Although SmartAid was developed for the micro-financing industry, its process can be applied to other sectors of development. As David Roodman of the Center for Global Development explains it:

30 Ibid.

31 In Canada see, "Participatory Budgeting – Working Together, Making a Difference" (Toronto: Toronto Community Housing, n.d.), available at http://www.torontohousing.ca/participatory_budgeting. In Chicago, PB is initiated in the city's 49th ward. See Alderman Joe Moor, 49th Ward Participatory Budgeting Initiative <http://www.ward49.com/participatory-budgeting/>, accessed July 27, 2010.

32 Suresh Balakrishnan and Sita Sekhar, "Holding the State to Account, Citizens Voice Through Report Cards in Bangalore" (Bangalore: Public Affairs Centre, February 2004).

33 Examples of process-based evaluation can be found in Appendix C.

34 World Economic Forum, "Philanthropy Measures Up," Global Leaders Tomorrow, Benchmarking Philanthropy Report, 2003.

35 CGAP is an independent policy and research center dedicated to advancing financial access for the world's poor. Housed at the World Bank, CGAP evaluates micro-financing institutions and provides market intelligence, promotes standards, develops innovative solutions, and offers advisory services.

How do you know a high-quality aid agency when you see one? Is it one that learns from past projects, or that rewards employees for subtle blend of results and risk-taking? Measuring such characteristics is hard, especially when the agencies being measured fund your budget. But that is exactly what CGAP has done, in the realm of aid for microfinance, using extensive surveys and interviews.³⁶

As a result of a peer review that analyzed the systems, procedures, and incentives of 17 of the most important microfinance donors, CGAP developed the SmartAid framework. The framework includes five measurable elements for evaluating micro-financing institutions (MFIs). The analysts selected nine weighted indicators to measure the five key elements and to assign organizations an overall score.

Another valuable aspect of this approach is that it allows an organization to determine where it has a comparative advantage in its field. Understanding which management and process elements an organization scores well in allows it to focus on its strengths and its comparative advantage. Similarly, such evaluations can provide insight into which organizations should forge partnerships. Forming a partnership with an organization that has complementary strengths can increase the overall effectiveness of the two organizations.³⁷

The United Nations Capital Development Fund (UNCDF) developed a similar evaluation that evaluates the processes as an indirect way to measure the outcomes of their programs.³⁸ One of the missions of UNCDF is to develop inclusive finance programs. To evaluate these programs, it has developed the “inclusive evaluation matrix.” The matrix asks a series of questions about the management, design, and process of a program. For example, questions include, “To what extent are the financial service program providing appropriate services and opportunities to women?” To measure this, the following indicators could be used 1) the number of women in senior management positions, including the board, and 2) percentage of active clients who are women. For each question, the matrix provides a list of indicators that can be used to answer the questions, methods on how to collect the data, and sources for finding the information.

Appendix C provides examples of process evaluation methodologies that are used by international organizations. Many of these methods are less costly than evaluating project impacts and therefore, many of these methods could be adopted by domestic community development organizations.

36 CGAP, “What Is Smart Aid?” (Washington, DC, CGAP, n.d.), available at <http://www.cgap.org/p/site/c/template.rc/1.11.7956/>, accessed August 9, 2010.

37 Brigit Helms and Alexia Latortue, “Aid Effectiveness Initiative: Microfinance Donor Peer Reviews, Elements of Donor Effectiveness in Microfinance: Policy Implications” (Washington, DC: CGAP, April 2004).

38 United Nations Development Programme, “Annex 3. Evaluation Terms of Reference Template and Quality Standards,” in *Handbook on Planning, Monitoring and Evaluating for Development Results* (New York: United Nations, 2009), available at www.undp.org/evaluation/handbook/Annex3.html#62.

Evaluation Infrastructure

The final category of evaluation, underlying all the above categories, includes tools that improve the infrastructure of evaluations knowledge. They improve the standardization of evaluation methodology, promote the use of evaluations, and distribute evaluation techniques and results. This category is extremely important because evaluations are only as valuable as their ability to be used to improve either current or future development projects, and this cannot happen if the information is not shared with other practitioners. In addition, practitioners must be able to easily understand this information, which means that the data must be standardized for impact development projects to be understood, replicated, or improved. Such standardization also leads to improved transparency and credibility, which in turn attracts more funding. Organizations and foundations are attempting to create standards for the evaluation and measurement of social impacts. These organizations are leading the way in creating the infrastructure that will allow for the effective dissemination of evaluation research.

One example of a group of organizations working to create evaluation research infrastructure is the Global Impact Investing Network (GIIN) and its “Impact Reporting and Investment Standards” (IRIS) initiative. GIIN is a network of nonprofits that include organizations such as the Rockefeller Foundation, Acumen Fund, and B Lab that are dedicated to increasing the effectiveness of impact investing.³⁹ Impact investments attempt to have positive social or environmental impacts while generating profit. IRIS is creating a common framework for defining, tracking, and reporting the performance of impact capital. As a recent report notes:

The IRIS initiative will build on these sector-specific efforts to create a common language that will allow comparison and communication across the breadth of organizations that have social or environmental impact as a primary driver. A common language for measuring and reporting performance forms a basis for enabling infrastructure and leads to transparency and credibility.⁴⁰

Currently the IRIS framework consists of hundreds of social indicators that are defined and categorized into five parts:

1. The social performance standards framework and description;
2. Descriptor Indicators that focus on the organization’s mission, products, and services and the target markets that the organization intends to affect;
3. Financial Indicators that are expected to be commonly reported for all organizations regardless of size or operational mode;

³⁹ For more information, see Global Impact Investing Network, <http://www.thegiin.org/cgi-bin/iowa/home/index.html>.

⁴⁰ The Rockefeller Foundation, Acumen Fund and B Lab, ‘Impact Reporting and Investment Standards’ website: <http://iris-standards.org/>.

4. Operations Indicators that are applied to organizations depending on their operational type and where in the production process they function; and
5. Sector-specific indicators:
 - agriculture and artisanal indicators
 - energy, environment and water indicators
 - education indicators
 - community development financial indicators
 - healthcare indicators
 - microfinance indicators

Similarly, an online database that is increasing the evaluation infrastructure is the “Tools and Resources for Assessing Social Impact” (TRASI), created by the Foundation Center. TRASI contains a comprehensive listing of 150 approaches to measuring and analyzing social impact for programs and investments. A range of organizations have contributed to the database, including social investors, foundations, NGOs, and microfinance institutions. These approaches range from off-the-shelf tools and concrete methodologies to generalized best practices. These approaches to measuring social impacts have been classified into 18 categories.⁴¹

Applying International Methodologies to Domestic Community Development

The above section demonstrated how various evaluation techniques used in international development have greatly enhanced the effectiveness of certain development projects. Unfortunately, evaluations are rarely done for domestic community development projects, and therefore the impacts of these projects are largely unknown. Using the examples from above, I suggest various ways international development evaluation methods could be adopted domestically to improve the cost-effectiveness of the billions of dollars that are spent each year on community development projects.

Financial institutions often fund community development projects in conjunction with nonprofits and federal, state, and local governments. These sources of funding can be categorized into five main groups:⁴²

Government: Federally, the U.S. Department of Treasury, through the Community Development Financial Institution (CDFI) Fund, channels financial support directly to the CDFIs that register with the fund in order to receive funding.

Depository Institutions: These include banks and credit unions that are created specifically for the purpose of working in markets underserved by traditional capital and federally insured commercial banks and savings institutions, which are motivated by CRA requirements.

41 The Foundation Center, “Tools and Resources for Assessing Social Impact” website at <http://trasi.foundationcenter.org>.

42 Ben Thornley and Colby Dailey, Building Scale in Community Impact Investing through Nonfinancial Performance Measurement,” Federal Reserve Bank of San Francisco, *Community Development Investment Review*, Volume 6, Issue 1, 2010.

Nondepository Institutions: These include institutions created specifically for the purpose of community investment, such as community development loan funds and community development venture capital funds registered with the CDFI Fund. They also include a diversity of pension funds, private insurers, financial advisors, and investment funds investing for financial and nonfinancial returns.

Philanthropic Organizations Making Investments: These include corporate, community, religious, and especially private foundations that invest in community development projects.

Individuals: Tens of thousands of individuals, including bank customers, mutual fund investors, and the world's wealthiest families, are a large contributor to community development projects.

A large portion of the money funneled to community development projects is motivated by the CRA.⁴³ Among the 998 institutions reporting CRA-motivated lending in 2007, 746 institutions invested \$63.8 billion.⁴⁴ None of these institutions are required to conduct evaluations of their investments.⁴⁵ For example, the single largest investor in community development projects in the United State, the CDFI Fund, requires only that CDFIs respond to a very narrow set of social impact survey questions. In any one year, just one-fifth of CDFIs are mandated to report community impacts.⁴⁶

As reported in their paper, *The Role of Nonfinancial Performance in Community Impact Investing*, Colby Dailey and Ben Thornley conducted a literature review, stakeholder interviews and examinations of 40 annual reports. They concluded that the reports lacked nonfinancial performance measurements in the field of community development. The nonfinancial returns were not widely measured nor widely disclosed. The review revealed significant differences and clear trends in the measurement and reporting of nonfinancial returns. The general observations they made were as follows:

- Most impact investors surveyed do not include nonfinancial performance in annual reports. The measures that are reported are usually published separately or only on the investor's website
- Although the CDFIs report on a very narrow set of survey questions, they report nonfinancial performance in the greatest depth compared to other community development practitioners, with measures of job creation, housing units and commercial/facilities spaces financed, number of individuals served, and minority group representation

43 Currently four federal regulators oversee the CRA: Federal Reserve Bank, Office of the Comptroller of the Currency, Federal Deposit Insurance Corporation, and the Office of Thrift Supervision. The latter will cease operation in the summer of 2011.

44 Federal Financial Institutions Examination Council, www.ffiec.gov, accessed July 21, 2010.

45 Ben Thornley and Colby Dailey, "Building Scale in Community Impact Investing through Nonfinancial Performance Measurement," Federal Reserve Bank of San Francisco, *Community Development Investment Review*, Volume 6, Issue 1, 2010.

46 Ibid.

- Banks, in particular, appear to use nonfinancial performance as a marketing and branding tool in annual reports, with stories and photographs but no accompanying analysis. All of the major banks sampled published separate corporate citizenship/corporate social responsibility (CSR) reports or reported on CRA lending volume on websites;
- Foundations, pension funds, and private equity managers were the least likely to report publicly on nonfinancial performance.
- Investment firms (fund/equity managers) generally highlight nonfinancial performance by describing screening and selection processes and the characteristics of underlying portfolio companies

The analysis also found that the most widely reported metrics were the volume of inputs invested for community impact (dollars of financing and capital); that anecdotal evidence of impact was typically presented as a supplement to quantitative performance measures; and that none of the investor annual reports reviewed benchmarked the nonfinancial performance data that was included.

Next I will use the template I created to look at how community development practices can be improved domestically.⁴⁷ The goal is to inspire innovation and to reveal potential opportunities for improving domestic evaluation practices. Specifically, I offer examples of how these methods can either be implemented for specific community development projects or incorporated into the existing institutional framework that is responsible for carrying out community development projects in the United States.

Randomized Evaluations

The scope of community development projects in the United States is large and diverse, but it is accepted that all could benefit from thorough evaluation. Affordable housing is a large sector of the community development field. Affordable housing typically refers to affordable rental housing for those of lesser means. Randomized evaluations could help uncover various impacts of stable and affordable housing on a family, which could improve the design of future affordable housing programs.

Recently, a *New York Times* article criticized the randomized evaluations of the affordable housing program Homebase as unethical and cruel.⁴⁸ However, there are usually more eligible candidates for affordable housing than available housing and funds. Therefore, a control group (those who did not receive housing) would exist irrespective of the randomized evaluations, eliminating ethical issues in randomly assigning eligible candidates to either receive affordable housing or not. This fact that there are more eligible candidates than funds

47 A more in-depth analysis of this topic can be found in Ben Thornley and Colby Dailey, "Building Scale in Community Impact Investing through Nonfinancial Performance Measurement," Federal Reserve Bank of San Francisco, Community Development Investment Review, Volume 6, Issue 1, 2010.

48 Cara Buckley, "To Test Housing Program, Some are Denied Aid," *New York Times*, December 8, 2010.

available provides an opportunity for a randomized evaluation as typically there are multiple treatment groups and a control group that can be evaluated and compared. A different community development intervention could be introduced to each treatment group to understand which aspects of affordable housing have the greatest impact on the tenants' livelihoods. Tenants who are selected would then be evaluated prior to moving into the housing and then re-evaluated periodically thereafter. These evaluations would measure social indicators such as housing conditions, health, employment, education, mobility, welfare receipt, delinquency, and so forth. These results would then be compared with the control group to reveal the impact of the program.

There have been very few randomized evaluations of community development projects. One example of such an evaluation is the Moving to Opportunity program. This on-going evaluation examined the short-term impacts of moving from high-poverty public housing to lower-poverty neighborhoods. Families that qualified received Section 8 housing vouchers through a random lottery. One group was required to move to low-income neighborhoods, a second group could use the vouchers to move anywhere, and a third group (the control) received no vouchers. The evaluators were able to determine whether moving to low-income neighborhoods caused changes to well-being relative to a control group.

The short-term results, which were measured five years after the treatment groups received the vouchers, showed that moving to low-poverty neighborhoods had significant positive impacts on personal safety, housing quality, mental health and obesity among adults, and staying in school, delinquency, and risky behavior among teenage girls. However, there was an apparent negative impact on boys' behavior and no significant effect on employment outcomes for adults or educational achievement for children. Only marginal improvements were found in the quality of schools attended.⁴⁹ These results are considered short-term impact and the effects of the program could change in the long-term.

A randomized evaluation such as this conducted for the HUD program rarely occurs for any community development project. Such randomized evaluations should be performed more frequently and continuously improved in order to clearly understand the effects of community development interventions and to determine any needed policy adjustments. It is not realistic to suggest that randomized evaluations should be performed on all, or even the majority, of community development projects. However, if they are performed on key projects, they can provide a tremendous amount of information for other projects and evaluations to build on.

Consider the deworming program in Kenya, for example. Without the proof of causality provided by the randomized evaluation, it would have been unlikely that developers would have considered allocating the majority of resources to a deworming program, as opposed

49 For interim results see Jeffrey Kling, Jeffrey B. Liebman, and Lawrence F. Katz. "Experimental Analysis of Neighborhood Effects," *Econometrica* 75(1) (2007): 83-119; and Jeffrey Kling, Jens Ludwig, and Lawrence F. Katz, "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment," *Quarterly Journal of Economics* 120(1) (2005): 87-130.

to additional teachers or materials, in order to increase school attendance. Randomized evaluations are needed to test our assumptions about what works and what does not work in community development, and, as the deworming example illustrates, to uncover seemingly unlikely solutions to achieve a development goal. In addition, randomized evaluations provide a solid base for other, less expensive evaluations, such as observational evaluations.

Community development practitioners should work together and select key projects for a randomized evaluation. These projects should be selected on their ability to allow for the testing of multiple interventions and assumptions and, at the same time, provide the most pertinent information for a large variety of community development practitioners. This way, a small number of randomized evaluations can provide a large amount of useful information to the community development field.

Observational Evaluations

As discussed above, randomized evaluations are not always feasible. Observational evaluations, however, might be a good second option. Hundreds of observational evaluations have been conducted on community development projects, largely performed by academic institutions. It is rare to find a community development policy or program that mandates an observational evaluation. One example is the U.S. Department of Education's "Promise Neighborhoods." The program is intended to significantly improve the educational and developmental outcomes of all children in the most distressed communities, including rural and tribal communities. The program aims to transform those communities by supporting comprehensive cradle-through-college-to-career programs.

Among 339 applicants, DOE only had funding for 20 distressed communities. An observational evaluation could provide valuable information on the effectiveness of the program. Data for each community that receives funding could be collected prior to receiving funding and periodically thereafter. This data could be compared among the 20 communities to see what interventions were implemented and which communities improved.⁵⁰

Although this would not demonstrate causality, it would provide useful information by revealing which interventions are most correlated with improvement in social indicators. There is a large spectrum of evaluations techniques that can be adapted to every type of community development project. It is just a matter of finding the right one that fits the specific characteristics and budget of a particular project.

If more government policies and community development projects required data collection and evaluation, the pool of data on these types of projects would grow. With better data, a greater number of, and more accurate, regression analyses could be conducted to reveal what works and what doesn't in community development. Paul Collier in *The Bottom Billion*, offers an example of how regression analyses using large databases has resulted in

⁵⁰ U.S. Department of Education, Promise Neighborhoods, <http://www.ed.gov/blog/2010/04/promise-neighborhoods/>, accessed July 27, 2010.

surprising and useful information. Collier used large databases from a number of sources, such as the U.N. and World Bank, to identify four main poverty traps for countries: 1) *conflict trap*—the more recently a country has experienced a civil war, the more likely it suffers from a lack of development; 2) *natural resource trap*—the more a country’s economy relies on one natural resource, the greater the likelihood of lacking development; 3) *landlocked trap*—landlocked countries are more likely to lack development; and 4) *bad government trap*—countries that have high corruption, patronage, and relatively high military spending are correlated with lack of development. This information is very helpful for development practitioners in designing new projects that address these poverty traps.⁵¹

Cost-Benefit Analysis

Government assessments of projects and programs rarely include a comprehensive cost-benefit analysis. Only about 12 percent of all the U.S. government regulations that had an impact assessment in 2007 included a cost-benefit analysis or a best estimate for total net benefits to society. This 12 percent does not include the many more regulations that had no impact assessment conducted.

Table 2. Summary of U.S. Regulatory Impact Analyses and EU Impact Assessments

<i>Scorecard item</i>	<i>Percent of analyses in U.S. study that include the scorecard item (n = 74)</i>	<i>Percent of analyses in European study that include the scorecard item (n = 70)</i>
<i>Estimation of total costs</i>		
Provided best estimate of total costs	65%	19%
Provided range of total costs	34%	13%
<i>Estimation of total benefits</i>		
Provided best estimate of total benefits	22%	13%
Provided range of total benefits	26%	3%
<i>Estimation of net benefits</i>		
Provided a best estimate of net benefits	12%	13%
Provided a range of net benefits	20%	4%

Notes: U.S. study figures taken from Hahn and Dudley (2007), based on regulatory impact analyses. European Study figures taken from Renda (2006), based on impact assessments. See text for details. Numbers are rounded to nearest percent.

An example of the tremendous amount of useful information a cost-benefit analysis of a community development program provides is a study on early care and education in the United States.⁵² This study used both randomized and observational evaluations and

51 Paul Collier, *The Bottom Billion*, Oxford University Press, 2007.

52 James Heckman, Rob Grunewald, and Arthur Reynolds, “The Dollars and Cents of Investing Early: Cost-Benefit Analysis in Early Care and Education, Zero to Three,” v26 n6 p10-17 July, 2006.

found that, compared with a control group who did not receive early child education, those who did were more likely to make significant gains in cognition, social-emotional development, to perform better in school, to have higher earnings in the workforce, and they were less likely to commit juvenile and adult crimes. The researchers also calculated the estimated economic benefit of the early care and education programs in increased earnings (and resulting tax revenues), public savings due to reduced crime, averted crime victim costs, and reduced need for rehabilitation and treatment. The range of the benefit to society was from \$3.78 per dollar invested to \$17.07 per dollar.⁵³ This type of analysis provides insight into which program should be further funded and expanded.⁵⁴

It is important that such cost-benefit analyses be encouraged, if not required, for projects implemented to fulfill CRA requirements and that standards are created and used. Standards—in this case that the assumptions (discussed above) for similar projects be calculated using the same methods—are important so that they may be accurately compared to one another. The greater the standardization, the easier it is to compare the benefits and costs of community development projects, even those that are quite dissimilar.

As mentioned above, randomized evaluation allowed development practitioners to illustrate that out of all the programs analyzed, the deworming program was the most effective program at increasing student attendance. In addition, a cost-benefit analysis was conducted and revealed the deworming program would cost on average \$3.50 per pupil per year. This was compared to the next most effective program, which cost approximately \$99 dollars per pupil per year. In addition, the evaluation estimated that the program cost only \$5 per Disability Adjusted Life Year (DALY)⁵⁵ saved. This is extremely low when compared to vaccinations that are considered one of the most cost effective ways to save DALYs at \$12 to \$17. With this information it is obvious that the more effective and cheaper deworming program is the one that should be scaled up.⁵⁶ Imagine how much money could be saved and how more effective community development projects would be if similar information could be obtained through evaluations.

Citizen Participation

Some have argued that participatory budgeting (PB) and community report cards (CRC) are not necessary in the United States because the democratic political system allows citizens to participate through elections. However, this is not always the case in high-poverty areas. A study conducted by the Federal Reserve System Community Affairs Offices found that local

53 This disparity in benefit-cost ratio is partially due to the varying time frames that were used to evaluate each program. The longer the time period was from implementation, the greater the benefits.

54 James Heckman, Rob Grunewald, and Arthur Reynolds, "The Dollars and Cents of Investing Early: Cost-Benefit Analysis in Early Care and Education, Zero to Three," v26 n6 p10-17 July, 2006.

55 The disability-adjusted life year (DALY) is a measure of overall disease burden, expressed as the number of years lost due to ill-health, disability or early death.

56 Edward Miguel and Michael Kremer, Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities, *Econometrica*, Vol. 72, No. 1, January, 2004.

governments are not always leaders in alleviating poverty in low-income areas.⁵⁷ This suggests that locally elected governments in these low-income areas are not adequately addressing the community development needs of the citizens. This situation could be improved if the citizens had more say in where funds were allocated.

Currently, it is often the case in the community development field that organizations or government agencies identify the needs of low-income communities and then procure funding to address those needs. Yet often the organizations assessing the needs of a community are not located in the community and do not take the time to gather input from those living there. Consequently, they may not be in the best position to determine the needs of that community. This situation is further exacerbated when multiple government agencies, nonprofits, and foundations operate in the same area and all compete for limited resources.

PB and CRC could address these problems by providing citizens with the opportunity to identify and prioritize those aspects of their community they wish to improve. This would push competing organizations to collaborate and share funding in order to meet the needs of the community as determined by those living in it.

One argument against greater citizen participation in the decision-making process is that bureaucracies and the political processes are already slow-moving and that providing additional avenues for citizen participation will bog down the political processes even further. An additional argument is that the citizens in these low-income communities usually have low levels of education and therefore will make less-informed decisions than practitioners in the community development field. However, these obstacles are even more prevalent and extreme in low-income areas of developing countries, yet PB and CRC have succeeded in improving communities in these countries. As mentioned above, PB practices in Brazil resulted in greater infrastructure improvements and a larger percent of resources allocated to health and education services. CRC in India resulted in a decline in corruption and an increase in citizen satisfaction with public services.⁵⁸

PB and CRC should be performed prior to implementing a community development project to gain insights on whether the particular project is a priority for the community or if they would prefer to allocate funds to a different community project. The information gained from PB and CRC would reduce the number of failed projects that are implemented, but never properly supported or used by the local community. In addition, such information could be analyzed for correlations between certain types of communities and the types of project and improvements they prefer. This will assist community development practitioners in designing projects that more directly address the needs of the community.

57 The Federal Reserve System and Brookings Institute, "The Enduring Challenge of Concentrated Poverty in America: Case Studies from Communities Across the U.S.," <http://www.frbfsf.org/cpreport/>, 2006.

58 Anwar Shah, ed., *Public Governance and Accountability Series, Participatory Budgeting* (New York: World Bank, June 9, 2010). See also Balakrishnan and Sekhar, "Citizens Voice Through Report Cards in Bangalore."

Process Evaluations

A further suggestion is to create a rating or certification system for banks, based on the comprehensiveness of their evaluations of community development projects. The certification process could be modeled after the Fair Trade Certification system and might work alongside, or even replace, current CRA regulation.

The Federal Reserve System, Office of the Comptroller of the Currency, or the Federal Deposit Insurance Corporation could conduct the certification process, or it could be contracted out to a third party. Banks could earn certification by allocating a certain percentage of their investments to certified community development projects or programs, funds, or intermediaries, such as CDFIs or Community Development Corporations (CDCs).⁵⁹ These third parties would be certified through conducting thorough evaluations of their activities. Banks would seek certifications as they could be used as a marketing tool and as part of their public relations strategy. A bank could attract customers with a certification demonstrating the bank's commitment to community development. The Fair Trade Certification has already proved there is a large market for those who want to purchase goods or services from organizations that contribute to sustainable development or environmental protection (see Appendix C for more information about Fair Trade Certification). Similar certificates include LEED certification for green buildings. In addition to a certification system, an added incentive could be created for those banks that demonstrate effectiveness in their community development projects. A bank could satisfy its CRA requirements by investing in CRA-certified projects or CRA-certified intermediaries, for example.

As an alternative to a certification system, a ranking or scoring system could be developed to rank those banks that have the greatest community development impact. Such ranking would create competition among banks to attract more socially conscious customers. Similar ranking and scoring systems have been developed for companies and nonprofits that are involved in social impact work.

A certification system could save both the banks and the bank regulators time and resources. Currently banks employ CRA compliance staff who are responsible for collecting and analyzing data on the bank's operations in low-income areas in order to demonstrate compliance. If these employees were redirected to evaluating community investment projects working toward social impacts the bank is supporting, then the bank could receive certification or a ranking based on established criteria. Certified banks could be rewarded, for example, by being subjected to less frequent CRA reviews. This would allow the banks to reduce the number of CRA compliance officers employed. In addition, they could have their CRA requirements reduced (as suggested with a certification system) based on their ranking or score.

⁵⁹ Although there is no established legal definition of CDCs, they are characterized by their community-based leadership and their work primarily in housing production and/or job creation. This is what differentiates them from other types of nonprofit groups.

Bank regulators also invest resources in bank examiners who are responsible for assessing whether a bank is complying with CRA requirements. Currently, CRA compliance staff at banks are required to produce large amounts of documentation to demonstrate that the bank activities meet CRA requirements. Bank examiners then must examine this documentation to assess whether each investment complies with CRA requirements. This process is time-consuming and expensive. Alternatively, if a bank can simply show that it invested in a certified CDFI, then a bank regulator would not need to review the investment because the CDFIs, by definition, comply with federal CRA requirements. However, there are not always CDFIs available in a bank's area of operation. If more individual projects/programs, CDFI Funds, or intermediaries, such as CDFIs and CDCs, were certified as CRA compliant, banks could then choose to invest in such entities, thus reducing the costs of the bank's CRA officers and the bank regulators' CRA compliance monitoring costs.

It is true that by reducing CRA requirements for banks that demonstrate effective investments in community development projects, these banks might invest less community development money because the amount of funds they will be obligated to invest in community development projects would be reduced and redirected to more profitable investments. However, the incentive would allow banks to discover those projects that are most effective and to allocate all of CRA funds to projects that are proven to be successful rather than to many projects whose impacts are undetermined. In addition, the information and data from such evaluations could be disseminated to other banks, which could then invest in similar community development projects. These benefits would far outweigh the negative impact of having CRA requirements reduced for banks that demonstrate effective community development investing.

A CDFI Fund certification process currently exists through the Financial Assistance (FA) Awards and the Bank Enterprise Awards Program (BEA). Through the FA Awards, the CDFI Fund invests in certified CDFIs that have demonstrated financial and managerial capacity to provide financial products that will positively impact their communities. Similarly, the BEA supports FDIC-insured financial institutions that are dedicated to financing and supporting community and economic development activities.⁶⁰ These certification processes can be

⁶⁰ The FA Awards are given to financial institutions that can demonstrate the ability to: 1) provide affordable and appropriate financial products and services that positively impact their communities; 2) be viable financial institutions; and 3) use and leverage CDFI Fund dollars effectively. The BEA Program provides financial incentives to those banks that help to expand investments in CDFIs and to increase lending, investment, and service activities within economically distressed communities. The BEA Program provides grants to applicants in three categories: 1) investments or technical assistance to qualified CDFI partners; 2) affordable mortgages, small business, education, commercial real estate loans; and 3) deposits, community services, and financial services. Certification and scoring appears to be the method that could most successfully be implemented on a large scale for community development projects in the United States. With existing community development institutions, there is considerable potential for including a thorough certification system for those banks and organizations that, through evaluations, can demonstrate effective allocation of resources to development projects. However, the certification should be based more on actual results determined by evaluations and less on processes and dollar amounts invested. Once banks have the incentive to demonstrate the effective allocation of their resources to community development projects the certification process can be substantially scaled up and improved from their present the standard of practice.

improved and scaled up, which would give organizations more incentive to conduct evaluations that would prove their eligibility for such certification programs.

Evaluation Infrastructure

It is paramount to promote evaluation infrastructure that will allow community development practitioners to share more information and experiences. Much of the information in IRIS and TRASI is applicable to domestic community development projects. However, a database specifically designed for community development projects in the United States would make it easier for community development practitioners to analyze and make decisions based on domestically collected data.

Development projects are extremely sensitive to the environment in which they are being implemented. Development projects in a developing country must be designed with a completely different set of goals, assumptions, cultural considerations, demographics, infrastructure, and institutions than domestic community development projects. In addition, the outcome of a particular development project will vary significantly depending on its location. Therefore, it is important to develop evaluation infrastructure that focuses specifically on domestic community development projects.⁶¹

For an evaluation infrastructure to be built, incentives are needed for practitioners to both evaluate projects and to disseminate the data. One way would be to simply make it a requirement. The government could require organizations to provide evaluation for all federally funded programs and then publish the findings. Alternatively, the government could create a market for evaluation data. The government could purchase evaluation data and project design information for community development projects that have proved successful. Similarly, if incentives were created for banks to have a greater community impact through methods discussed above, such as a certification process or reduced CRA requirements, the banks would be willing to pay for the project design and evaluation information of successful community development projects. As illustrated above, a cost-benefit analysis on projects striving to address certain development issues can reveal large differences in their efficiency. If the impact of a bank's community reinvestment programs were important to the bank, it would appreciate the savings of a program that generates \$17 of benefit per \$1 spent as opposed to \$3. This potential savings would induce banks to pay for the information

61 For example, when analyzing evaluation methods of international development projects for the purpose of applying them to community development, two obvious differences must be considered. First, evaluation methods for international projects were usually developed to assess populations living at poverty levels rarely seen in the United States. It is also unlikely that U.S. residents lack access to clean water, medical facilities, basic infrastructure and institutions. Methods must either be adapted for the United States environment or abandoned as incompatible. Second, discrimination based on gender is much greater in developing countries. As a result, it is important to measure the social impacts by gender in international development projects. In the United States, discrimination is most evident against minorities. Again, this must be considered when adopting international evaluation methods in the United States. These differences do, however, provide an opportunity to adopt innovative techniques for measuring gender discrimination to race-ethnicity discrimination.

that would allow them to implement the more efficient development program.

An evaluation information market would attract other parties besides the government or community development practitioners. Private consultants and specialists who focus on performing evaluations of community development programs would perform evaluations and gather data in return for a fee. As cost-benefit illustrated, there is the potential for significant efficiency improvements in community development programs—perhaps enough so that such consulting fees could be covered.

The U.S. government is beginning to recognize the importance of evaluating development and other social impact programs. In May 2005, the Panel on the Nonprofit Sector (established by Independent Sector) recommended to the Senate Finance Committee that, as a best practice, charitable organizations should establish procedures for measuring and evaluating their program accomplishments on the basis of specific goals and objectives.⁶² In addition, the panel recommended a sectorwide effort to provide information and training focused on appropriate methods of evaluations of social impact programs. Although it appears improbable that detailed federal legislation on performance reporting will be developed, practitioners are focusing more attention on evaluation processes in order to assess their effectiveness.⁶³

In 2009, President Obama's White House Office of Social Innovation and Civic Participation asked Congress to provide \$50 million in seed capital to identify the most promising, results-oriented (determined through evaluations) nonprofit programs and expand their reach throughout the country. The guidelines include the following three principles: 1) share knowledge across stakeholders within and outside government; 2) learn about the best solutions in communities; and 3) continually update principles, standards, and expectations of measurement and evaluation.

Similarly, the U.S. Agency for International Aid (USAID) has its own evaluation requirements for the development projects it invests in overseas.⁶⁴ These evaluation requirements are listed in its Automated Directives System (ADS), chapter 203, "Assessing and Learning." Chapter 203 provides guidance for USAID operating units on agency practices and standards used to determine the attainment of project objectives. The guidelines address how to collect and present impact data; use evaluations to improve development outcomes; identify obstacles that are preventing intended results and how to overcome these obstacles, and disseminate evaluation results to improve the practices of the development community as a whole.⁶⁵

62 Linda M. Lampkin et al., "Building a Common Outcome Framework To Measure Nonprofit Performance" (Washington, DC: Urban Institute, 2007), available at www.urban.org/publications/411404.html.

63 Independent Sector is the leadership forum for charities, foundations, and corporate giving programs committed to advancing the common good in America and around the world. See also, Independent Sector, <http://www.independentsector.org/about>.

64 USAID is the principal U.S. agency to extend assistance to countries recovering from disaster, trying to escape poverty, and engaging in democratic reforms.

65 USAID, Automated Directives System Chapter 203, Assessing and Learning, April 2, 2010.

The USAID directives for evaluations and the work done by the Office of Social Innovation and Civic Participation on improving evaluation methodology should be used to develop a similar department that would assist those in the United States who invest in community development projects to improve their evaluations methodology and eventually their effectiveness in achieving social change.

Conclusion

Evaluating development projects is vital for improving the lives of those affected the most by poverty. Evaluations will allow stakeholders to identify the most effective community development projects, which can then be expanded for a greater impact. The United States should implement new policies that will inspire a revolution in evaluation practices.

The federal government must play a larger role in developing and applying evaluations for community development projects. The federal government is the largest single collector and repository of social impact information and investing.⁶⁶ Evaluation research should be considered a public good, like the construction of a freeway or other infrastructure projects, in that the total benefits from the research may far exceed those experienced by the people involved in the study.⁶⁷ Information from an evaluation can be used to design and implement more effective and efficient community development projects, which will produce greater social impacts and savings in government expenditures. Evaluation research loses most of its value if it is not shared with other community development practitioners. Like many public goods, the government must take the lead because most organizations will not invest in developing and sharing evaluation research because of a “free rider” problem (those who use the information do not contribute to the cost of collecting it).

The federal government also has the ability to standardize evaluation research and reporting procedures for community development projects. It would be helpful to require certain evaluation methods and reporting procedures in order for community development practitioners to receive tax breaks or other financial benefits.

The current CRA provides little incentive for banks to conduct evaluations of community development projects they invest in. As a result, little is known about the true impact of the CRA on low- and moderate-income communities in the United States. Banks do not want to allocate the additional money and time to collect and analyze data. Banks are simply trying to meet their CRA requirements and are not invested in the social impact of the projects. These incentives must be altered to encourage banks to perform evaluations and to invest in more effective community development projects.

66 Ben Thornley and Colby Dailey, “Building Scale in Community Impact Investing through Nonfinancial Performance Measurement,” Federal Reserve Bank of San Francisco, *Community Development Investment Review*, Volume 6, Issue 1, 2010.

67 Franck S. Wiebe, Millennium Challenge Corporation, “Working Paper: Aid Effectiveness: Putting Results at the Forefront,” October 27, 2008.

It is time to create an evaluation environment in the United States that creates incentives for thorough evaluations and shared impact results. The international arena offers a wealth of knowledge that could inform and inspire domestic antipoverty efforts. With more and better social impact evaluations and cost-benefit analyses, we can have a greater impact with fewer resources on U.S. impoverished communities. We must demand proof of the effectiveness of our community development projects and no longer be content with basing our community development strategies on status quo assumptions.

Paul Veldman is currently completing a Master's Degree in Public Administration (MPA) at Columbia University's School of International and Public Affairs (SIPA). He also has experience working with microfinancing institutions in Kenya and currently works with the United Nations Population Fund (UNFPA). Prior to attending SIPA, he worked as an Environmental Impact Consultant in San Francisco. Mr. Veldman holds a B.A. from UC Berkeley.

Appendix A: Examples of Observational Method

The following are examples of observational evaluation methods that are used in the international development field. Many of these examples use a mix of the above observational evaluation techniques.

Progress Out of Poverty (PPI)

Overview. Dr. Mark Schreiner developed the Progress Out of Poverty Indices (PPIs) for the Grameen Foundation. The PPIs attempt to measure how many microfinance clients are in poverty and whether they are moving out of poverty. Although this method has been applied to the microfinance sectors, in theory it could be applied to other industries. The method used in the PPIs is very similar to the method used by major credit card companies when determining the credit risk of potential clients, but adapted to estimate the likelihood of poverty. As of 2008, foundations and organizations have developed approximately 36 PPIs in the following countries: Bangladesh, Bolivia, Guatemala, Ghana, Haiti, India, Kenya, Malawi, Mexico, Morocco, Nepal, Nigeria, Pakistan, Peru, South Africa, The Philippines, and Vietnam.

Data. The PPI is based on a national income and expenditure household survey for a country. This usually involves a sample size of 7,000-50,000 households, which is usually conducted by a government's statistical department every several years.

Methodology. The survey asks questions of households that assess how much a household earns and how much it spends, as well as up to 200 other nonfinancial questions such as: What are the walls of your house made of? Do you own a refrigerator? How many in your household are between the ages of 0 and 17? Analysts then use regression analysis of the survey data to determine which of the nonfinancial questions best predict poverty at the national or international poverty line for a given country. The PPI ultimately consists of ten questions that are predictive of poverty, not co-related (overlapping), easy to ask and not offensive, easy to verify, and discreet. The questions are tailored to local customs and personnel preferences in different countries. Loan officers at micro-financing institution (MFI) ask the questions of clients during the loan application and maintenance process. This approach uses the prospect of receiving a loan as an incentive to beneficiaries to participate in the evaluation process.

Analysis. From the surveys, a threshold is calculated to represent the poverty line. Data are then aggregated to illustrate trends among groups of clients over time, and their progress out of poverty is subsequently measured. This approach only assesses indirect effects and, therefore, cannot be used to measure the MFIs' direct impact on poverty reduction.

Dalberg Approach

Overview. The Dalberg approach was created by the Dalberg Global Development Advisors in 2001 to assess organizations' progress in achieving financial and social goals. As of 2008, Dalberg has implemented more than 400 projects for more than 100 global organizations in over 75 countries around the world. About one-third of projects evaluated have been for international organizations such as the World Bank and the United Nations Development Programme (UNDP); one-third with businesses trying to expand operations into developing countries, and one-third with foundations such as the Bill and Melinda Gates Foundation. This approach provides a more accurate assessment of how an organization's social impact activities are benefiting an individual or community and focuses more on the indirect impacts of an organization's activities and less on its direct impacts.

Methodology. The Dalberg approach divides organizations involved in social entrepreneurial activities into three categories: 1) those striving for standard commercial returns with some consideration for social impacts; 2) those seeking blended capital models where a lower internal rate of return is acceptable in order to achieve social impacts but profitability is still expected; and 3) social-enterprises that may not be commercially viable, but are striving to achieve self-sustainability. Most of the data are self-reported by the client investor or company.

Data. This approach collects data on a program or project's result hierarchy, discussed above. Analysts collect information on the entire process from inputs, activities, outputs, outcomes, to goals.

Analysis. The Dalberg approach is a comprehensive analysis, starting with an evaluation of the social impact theory behind a project's activities. The approach then analyzes a project's inputs and the resulting activities from those inputs. Following this, researchers analyze the outputs of the project's activities and measure their resulting social impacts. Finally, the team compares these results against a counterfactual or benchmark, which is usually a similar business that focuses only on the bottom line and not a social impact. For example, the Dalberg approach would analyze an energy resource project that focuses not only on increased dollar volume purchased locally, but also on how to increase value added locally, and skills, employment, and enterprises built locally. This would be compared against a similar energy resources project that was concerned with just profits.⁶⁸

Movement Above the US \$1 a Day Threshold

Overview. The Movement Above the US \$1 a Day Threshold is a nine-year strategic plan begun in 1997 by the Microcredit Summit Campaign (MSC). The plan is to achieve its goal of reaching 175 million of the world's poorest people through microcredit and ensuring that, between 1990 and 2015, 100 million families rise above the US\$1/day threshold. The

⁶⁸ See www.dalberg.com/case_studies.php, for more information on the Dalberg approach.

movement is based on the partnership of MSC and 15 of the largest MFIs and networks in the world (representing approximately 60 percent of the world's entire microcredit clients in 2008). The MSC found sufficient existing data to make calculations for Bangladesh, India, Pakistan, Vietnam and Sri Lanka.⁶⁹ MSC plans to expand the project to include Africa and Latin America in the coming years. This approach only measures the economic status of the beneficiaries and focuses on the indirect impacts of the microfinancing programs and less on the direct impacts.

Data. The socioeconomic indicators used in the scorecard survey include information on food consumption, children's education, and various household assets. MFIs report the data, but credibility varies. MSC performs offsite third-party verification of source documentation by confirming the consistency of self-reported data.

Methodology. Assessing income improvement among clients includes: 1) analyzing existing data about clients' rise above the \$1/day threshold; 2) administering surveys to establish baseline data for new clients so their progress can be assessed; and 3) commissioning panels of top poverty researchers in various countries to ensure accurate estimates of the total number of clients who rise above the \$1/day threshold. In addition to the three steps, MFI loan officers administer a "poverty scorecard" of 10 questions during the loan application and maintenance interviews with the clients.

Analysis. The scorecard is administered at the clients' entry and periodically over time and hence determines whether, and when, a client passes the threshold. Partnering nongovernmental organizations administer the scorecards.

Social Rating

Overview. Social Rating is a tool pioneered in 2005 by Micro-Credit Ratings International Limited (M-CRIL).⁷⁰ The Social Rating has been applied to MFIs in Bangladesh, India, Vietnam, Cambodia, the Philippines, Kenya, South Africa, Bolivia, and Haiti. This approach provides more information about the direct effects of the MFI's activities and less on the indirect effects.

Data. Data are collected through client interviews targeting four areas: 1) clients' awareness about the financial projects, including knowledge of the interest rates they are charged on loans, the rate paid on their savings, and other details; 2) clients' access to capital and the role the MFI plays, such as whether anyone in the household already has a savings account with another MFI, has outstanding debt from a moneylender, etc.; 3) enterprise-level information including the enterprise's industry sector, whether any employees are non-family

69 www.microcreditsummit.org/movement_above_one_dollar_per_day/ for more information on the Microcredit Summit Campaign

70 M-CRIL provides a rating of MFIs that allows investors and to effectively invest their money to achieve social, ethical, and financial goals.

members, etc.; and 4) poverty information using, among other indicators, the PPI where applicable, and income relative to the poverty line where it is not.

Methodology. The process begins with an MFI providing M-CRIL with annual reports and other operational and portfolio information. M-CRIL reviews the reports, and an independent rating committee of third-party experts reviews the content and quality before the evaluation is finalized. After reviewing the material, M-CRIL staff meet with the MFI's board members and senior management to explain the Social Rating process. Analysts then interview a subset of staff members on all levels about mission, targeting, product development, market segmentation, client retention, human resources, management information system, and internal audit. Following this process, the team visits two to three branches for interviews with loan officers and a random sample of clients big enough to attain results at a 95 percent statistical confidence level. MFI staff who have been trained and who are supervised by M-CRIL often conduct the interviews.

Analysis. M-CRIL conducts client focus groups to obtain information on the effectiveness of the MFI. M-CRIL also performs cross-tabulation, frequency distribution, and qualitative analysis. The above efforts are used to assign a letter rating and accompanying narrative and quantitative report for the MFI. This information is made available to the public.

Appendix B: Examples of Cost-Benefit Analysis

Although cost-benefit analyses do not require as much time and resources as randomized evaluations, they still can be taxing on an organization. Collecting data for calculating the cost-benefit ratios can be difficult and time consuming. The following evaluation toolkits employ some form of a cost benefit analysis.

Acumen Fund, Best Available Charitable Option (BACO) Ratio

Overview. Acumen developed its BACO Ratio methodology in 2004. Acumen Fund seeks to quantify an investment's social impact and compare it with existing charitable options for that explicit social issue. This helps to inform investors where their philanthropic capital will be most effective and answers the question, "For each dollar invested, how much social output will this generate over the life of the investment relative to the best available charitable option?" In cases where a viable local comparison does not exist, Acumen tries to develop realistic hypothetical options on the basis of other geographies or from plausible "what if" scenarios.

Data. Acumen Fund assesses how another organization performs compared to outcomes Acumen Fund estimates they could achieve by using the same amount of money and implementing the program themselves. For example, Acumen Fund could estimate that it would cost them less than \$0.02 to protect one individual from malaria for one year, compared with \$0.84 via the next best option of investing the money with an NGO that already provides protection from malaria. In other words, Acumen Fund's investment in this scenario is 52 times more cost-effective than the best available charitable option.

Analysis. The BACO cost calculation is completed against a range of three financial scenarios: full return on investment (principal plus interest); return of only the principal; and complete loss. Similarly, the social impact forecasts are broken down into three scenarios: initial projections (from the original investment plan); conservative projections (developed by Acumen Fund portfolio team, on the basis of moderate growth plans); and revised projections (updated on a real-time basis using actual impact data).⁷¹

Hewlett Foundation Expected Return

Overview. The Hewlett Foundation developed its Expected Return (ER) methodology to evaluate potential charitable investments through a systematic, consistent, quantitative process.

71 Acumen Fund, The Best Available Charitable Option, http://www.acumenfund.org/uploads/assets/documents/BACO_percent20Concept_percent20Paper_percent20final_B1cNOVEM.pdf, January 2007.

Data. The formulae for the ER method is as follows:

Expected return is calculated as such:

$$\frac{\text{Benefit in perfect world} \times \text{likelihood of success} \times \text{philanthropy's contribution}}{\text{Cost}}$$

Where:

Philanthropy contribution includes a *financial contribution* (the percentage of an individual organization's contribution relative to the overall philanthropic contribution needed to achieve the outcome) and the *degree of influence* (how important the investment is to achieving the outcome). The degree of influence can result in philanthropic contribution that is greater than the level of financial share (if the philanthropy is providing other expertise, such as leadership), or it can result in a contribution that is less than financial share (if the philanthropy is relying on the services and expertise of others, and is contributing little other than money).

and where:

The cost includes program cost and percent of overhead cost used to support the program.

Likelihood of success is calculated as such:

$$\text{Likelihood of Success} = \text{strategic accuracy} \times \text{grantee success} \times \text{external conditions}$$

Analysis. Expected Return forces program officers to test their assumptions and theory of change or their logic model against the ER results, quantify high-level trade-offs between investments within an investment portfolio, and ideally make better prospective funding decisions within their investment portfolios. Calculating the likelihood of success forces program officers to consider unidentified risks instead of just focusing on impacts. Similarly, assumptions must be made explicit.⁷²

RAND Corporation

Overview. The Rand Corporation studied the measurement and use of estimated economic value, or "shadow price," in applying cost-benefit analysis to social program evaluation. Although this study is not an evaluation toolkit, its findings and method provide valuable insight on how to perform cost-benefit analyses.⁷³

Data. The RAND Corporation reviewed 39 social programs whose effectiveness has been evaluated with scientifically rigorous methods. A third party then conducted one or more cost-benefit analyses for 22 of the 39 programs.

72 William and Flora Hewlett Foundation, "Making Every Dollar Count: How expected return can transform philanthropy," <http://www.hewlett.org/what-we-re-learning/strategy>, April 10, 2008.

73 Lyn Karoly, "Valuing Benefits in Benefit-Cost Studies of Social Programs," Rand Corporation, http://www.rand.org/pubs/technical_reports/2008/RAND_TR643.pdf, 2008.

Analysis. The study finds that many of the important benefits that accrue from social programs are rarely, if ever, monetized. It also finds that shadow prices for outcomes in cost-benefit analyses do not consistently capture the full range of societal benefits or costs. Even where there is a well-established research base for valuing outcomes, the use of shadow prices is not consistent across studies of social programs. The study also found that program benefits that extend into the future may be monetized, but uncertainty associated with future costs and benefits must be recognized. The findings suggest directions for future methodological work to advance the use of cost-benefit analysis in evaluating the economic returns from social programs.

The Roberts Enterprise Development Fund SROI

Overview. The Roberts Enterprise Development Fund (REDF) created the Social Return on Investment (SROI) in 1996 to analyze the social returns on investment of two of its social purpose enterprises. Although the SROI was developed in the United States to evaluate domestic programs, REDF was one of the pioneering organizations to develop a cost-benefit analysis method for development programs. As a result, many international organizations have adopted the SROI or developed a new method based on it. The SROI was calculated to demonstrate the social, enterprise, and blended value accrued to society compared with the total investments for a given social enterprise program on an ongoing and retrospective basis.

Data. Data are mostly collected through analyses of the financial statements of the participating organization.

Analysis. The SROI measures value creation, which is presented as a continuum ranging from social, to socioeconomic, to economic value creation. The SROI uses a discounted cash flow analysis in an effort to monetize the economic value of social impacts of programs in the REDF portfolio. This monetized social value is then consolidated with the economic value created by the same social purpose enterprises.

The six stages of the SROI analysis are:

Stage 1: Calculate enterprise value, forecast its cash flow, and then discount its free cash flows, using an appropriate discount rate.

Stage 2: Calculate the social purpose value (mirroring the enterprise valuation process). Estimate the future social purpose cash flows, calculate a terminal value (value created beyond the ten-year point), and then discount that value using a discount rate appropriate to social investments.

Stage 3: Calculate the blended value: Add together the enterprise value and the social purpose value, and subtract any accrued long-term debt.

Stage 4: Enterprise index of return: This index summarizes an enterprise's financial performance relative to the investment it required. To calculate, divide the enterprise value (its projected cash flows discounted to a present value) by the investment to date.

Stage 5: Social purpose index of return: This index summarizes the degree of social impact that can be monetized relative to the investment it required. To calculate, divide the social purpose value (determined by the methods described above) by the program's investment to date.

Stage 6: Calculate blended index of return by comparing the blended value and the investment to date. This shows the return on both business and social mission activities relative to the investment to date.⁷⁴

Social Impact Assessment (SIA)

Overview. The SIA is an evaluation tool developed by Global Social Venture Competition (GSVC)⁷⁵ with methods adopted from REDF's SROI and the work of Clark et al. in the Double Bottom line Catalog. GSVC started requiring quantification of projected social impacts during the first round of the competition in 1999-2000, and formalized the SIA guidelines in 2003. As of 2008, 37 countries had submitted 764 full business plans, all of which contained SIAs.

Data. The SIA is based on three major steps: 1) define the social value proposition that is core to the organization's desired social outcomes using the "Theory of Change" tool (meaning that the proposition is strongly based on social impact theory); 2) quantify the social value by listing the top three leading social indicators that are measurable and that will most strongly correlate with the organization's desired social outcomes, and that can be tracked as part of their normal business operations; and 3) monetize the value of the social impact the organization aims to create over the next 10 years. Data are self-reported, and no one verifies the secondary research sources upon which extrapolated outcomes are based.

Analysis. The SIA provides guidelines for a self-directed process for entrepreneurs. The guidelines include developing an impact value chain that specifies financial, human, and other inputs required for operations; activities; measurable outputs produced; and outcomes or changes in terms of the social, environmental, or economic issues being addressed by the venture. Entrepreneurs prioritize outcomes and determine applicable leading indicators that are used to monitor the activities and/or outputs believed to be correlated with desired outcomes. Secondary measures of outcomes that have been correlated with similar ventures support the selection of metrics. If possible, the SIA directs entrepreneurs to assign a monetary value to outcomes, which allows for a discounted "social cash flow analysis" of these values. There are no standard discount rates for such calculations and, therefore, the SIA requires the entrepreneurs to create their own logical discount rates. The SIA is limited in scope because the evaluation is only for future projects and is not used for follow-up verification.

74 Carla I. Javits, "Social Return on Investment (SROI) Collection" (San Francisco: Roberts Enterprise Development Fund, n.d.), available at www.redf.org/learn-from-redf/publications/119, 2000.

75 The GSVC is the largest and oldest student-led business plan competition providing mentoring, exposure and prizes for social ventures from around the world. For more information, see www.gsvc.org/about_gsvc/.

Appendix C: Examples of Processed-based Evaluation

Evaluation of an organization's processes and management lends itself to creating a certification or scoring system based on how well organizations achieve key elements (that is, internal management systems, processes and procedures, and incentives) that have proven to result in successful development projects. Below is a discussion and examples of certification and scoring methods that use process-based evaluations.

Certification. The Fair Trade certification creates the market for impoverished farmers to sell their goods to the consumers with the assurance that the goods were produced in a socially responsible and environmentally sustainable fashion. Companies such as Starbucks have realized the benefits of attracting socially conscious consumers and now advertise that they are Fair Trade certified.

Fair Trade Certification

Overview. Transfair USA has developed a successful certification system for agricultural products grown in other countries. It began in the 1980s, and as of 2008 has certified more than 25 products from over 55 countries. Fair Trade Labeling Organizations International (FLO) certifies and updates certifications. FLO is an organization composed of nonprofit organizations and regional farmers representing approximately 1.4 million FT farmers and workers. The certification provides an evaluation of an agricultural producer's direct and indirect impact on economic social and environmental factors.

Data. A cooperative or producer group interested in receiving the FT certification submits an application. FLO-CERT, a wholly owned subsidiary of FLO, inspects participating farms and registered producers annually or biannually. FLO-CERT also verifies that the pricing premiums generated are invested in socially responsible projects, such as in community development projects, infrastructure, scholarships, or other needs.

Analysis. The certification guarantees to retail buyers that the agricultural products are produced with methods that meet the minimum standards for fair prices, fair labor conditions, direct trade, democratic and transparent organization, community development, and environmental sustainability.

Scoring and rating. Consumers and investors are becoming more socially conscious and as a result are seeking businesses that are socially responsible and that practice environmentally and developmentally sustainable practices. This has resulted in businesses focusing on a "double bottom line." The conventional first bottom line measures a business's fiscal performance—whether it is making a profit. The second bottom line assesses and measures its positive social impact. In addition, some businesses monitor their triple bottom line, which assesses the ecological impact.

Similar to the certification process, scoring and ranking businesses based on their social

impact are other possibilities. Rating an organization's performance can also be achieved with "benchmarking." Benchmarking evaluates an organization's performance relative to its peers. Benchmarks might be based on beneficiaries' perceptions of an organization's philanthropic activities, for example. If enough organizations are surveyed a ranking system can be created by comparing the performances, as determined by the surveys, of the organizations.

The various methods score organizations on a variety of aspects including: governance, impact on the company's employees, the community, the environment and consumers social impact, internal processes, learning and growth, financial risks, economy, society, well-being, and synergy. The following are examples of evaluations toolkits that are used to measure and rate organizations based on the various social impacts, and in some cases environmental sustainability, they achieve.

Human Impact + Profit (HIPTM) Scorecard and Framework

Overview. HIP Investor, Inc., developed the HIP Scorecard and Framework in 2004. It borrows from the SROI framework, among others. Analysts have used HIP to evaluate organizations in Costa Rica, France, Italy, India, The Netherlands, Thailand, United Kingdom, and United States. As of 2008, it has been applied to approximately 60 companies in the following sectors: energy, banking and microfinance, consumer projects and food, high technology, real estate, manufacturing, clean technology, and nonprofit organizations.

Data. The HIP quantifies human, social, and environmental impacts, how those impacts create financial results, and the management systems required to maintain success in the future. This approach focuses on the result-oriented measures of five categories: health (physical and mental), wealth (net assets and income), earth (carbon and environmental), equality (gender and ethnic balance), and trust (lawfulness and transparency). It assesses both direct and indirect impacts by analyzing the customers, employees, and suppliers, and it illustrates to what extent improvements in human impact create higher revenue, lower costs, or tax benefits. These measures are quantified and then an analysis is conducted through company interviews and secondary research. The HIP also assesses five management practices that drive sustainable, profitable growth: vision, measurement, decision making, accountability, and financial alignment.

Scoring. Analysts score organizations on the basis of three aspects: human impact, profit (and how it is affected by the human impact), and management practices (and the organization's relationship to sustainability). The scorecards can then be used for comparisons and to gauge the attractiveness and weightings of investments in a portfolio.

Portfolio Data Management System (PDMS)

Overview. The Acumen Fund in conjunction with Google engineers developed the PDMS in 2005. As of 2008, Acumen has used the system to assess more than 20 companies and more than 40 corporate and private foundations in countries including the United States,

Kenya, Pakistan, South Africa, India, and Tanzania. The PDMS is an online tool that allows investors to track a consistent set of quantitative financial, operational, and social metrics for each company. The PDMS primary focus is on the indirect effects of a company's activities, with less emphasis on its direct effects.

Data. Acumen staff enter company-reported data into the PDMS. In some cases, staff verify data on-site. Acumen creates a profile of a company in the PDMS when it is added to its portfolio, and sets targets on the basis of the company's business plan. The PDMS provides the flexibility to define custom metrics across investments and to perform qualitative assessments of the data. The impact potential of a company is based in part on a comparison of the company's outcomes per dollar invested with Acumen's "best available charitable option" (BACO), discussed above.

Scoring. The PDMS provides a qualitative rating of a company's management using a standardized "capabilities assessment" in six areas: alignment with the investors' mission, financial sustainability, potential for scaling up, potential for social impact, management capability, and business model effectiveness. This information is for investor use only and is not available to the public.

Charity Analysis Tool (CHAT)

Overview. New Philanthropy Capital (NPC), a U.K. organization, created CHAT in 2002. CHAT has been updated several times to improve its consistency of application and standardization of documentation.⁷⁶ This evaluation method provides a more in-depth analysis of an organization's indirect impact relative to its direct impact. CHAT has been used to assess 400 to 500 U.K. charities. NPC has performed partial analyses of organizations operating in other countries but does not officially list them as "NPC recommended."

Data. NPC analysts spend 9 to 12 months performing field and desk research and analysis of a specific sector in order to produce a list of 10 organizations that are published in the "NPC recommend" list. An important feature of the CHAT is that it subtracts base case evidence (taking into account what would have happened in the absence of the organizations' interventions) from the results when this information is available. This provides a more accurate measure of an organization's impact.

Scoring. The analysis rates organizations in three categories: results, risks, and capacity. The results section is further broken down into breadth, depth, and change. The risk section is divided into organizational, financial, management, strategy, and evidence of claimed impact. The ratings are summarized in a two-page report that scores the following topics on a scale of 1 to 5: breadth (scale), depth (intensity), and change of impact, risk, difficulty of fundraising, organizational maturity, innovation, scalability, replicability, and geography.

⁷⁶ NPC provides consulting to funders and nonprofits for the purpose of achieving greater impacts.

Compass Investment Sustainability Assessment

Overview. AtKisson, Inc., developed the Compass Investment Sustainability Assessment in 2000 to allow investors to choose companies that can effectively make the transition to achieving social, economic, and environmental sustainability. As of 2008, it has been used to evaluate 13 companies. A streamlined version has been used to assess approximately 75 companies using publicly available data. The companies span France, Germany, Indonesia, Japan, Sweden, United Kingdom, United States, and Thailand.

Data. AtKisson performs the analysis in conjunction with the target company, and it is used to identify the opportunity for the company to improve its score over the lifetime of the investment.⁷⁷ This assessment is an index evaluation of early-stage companies. It assesses progress in five categories: nature, economy, society, well-being, and synergy (how the company and all five categories complement one another to achieve its goals). These five categories are further divided into 20 parameters that are weighted according to the company's focus and area of operation.

Scoring. The analysis results in a point score on a 100-point scale. The following metrics are qualitatively assessed to estimate their direct and indirect impacts: energy usage, material flows, and interaction with the community. The reports are presented to each company and its investors. The assessment is updated every two to three years.

Development Tracking System (DOTS)

Overview. The International Finance Corporation (IFC) developed DOTS in 2005 to help evaluate and improve the social impact performance of organizations in its portfolio. IFC intends to use DOTS to track all of its projects, which are located on every continent. DOTS provides an annual development outcome rating and industry-specific social impact indicators that measure the social impact of investments on stakeholders. The result is a rating of an organization's indirect impacts and a partial analysis of its direct impact.

Data. A team of IFC employees identifies specific goals and performance indicators that apply to a specific project. The indicators must be relevant, able to be aggregated, time-bound, and easy to track. The team measures these indicators against benchmarks and timelines at least once.

Scoring. The analysts rate each project on four categories: financial, economic, environmental and social performance, and private-sector development impacts. The metrics together measure whether the investment has generated additional investments in the area or sector. Analysts assign the performance on these four categories and their interrelation a rating on a six-point scale from highly successful to highly unsuccessful. Results from DOTS are available to organizations that partner with its Advisory Serviced Division but are not publically available.

⁷⁷ AtKisson, Inc., is a consultant company that provides strategic consulting, training and facilitation, assessment, communications and reports, and workshops for organizations that are striving to accomplish sustainable development. <http://www.atkisson.com/wwd.php>.