

Firms Doing Good: How Do We Know?

Measurement of Social and Environmental Results

Michael Klein

Sumeet Kaur

The World Bank
International Finance Corporation
Inclusive Business Department
February 2014



Abstract

Social impact investors, philanthropists, or corporations pursuing social responsibility try to demonstrate that they are indeed “doing good.” This essay classifies the various types of measures that currently exist to capture social and environmental impact in a simple scheme. It argues that there is a basic “staircase of results measurement.” A first level of measures captures some aspect of “organizational readiness.” The next level describes some form of “result” that may or may not be attributable to the organization trying to do good. The third level gets at “impact” that can be attributed to an intervention. Beyond this, there are measures that assess the costs and benefits of interventions, allow aggregation of results from different interventions and comparison among them or across time. Finally, the essay discusses how measures are tied to incentives. It argues that the various approaches can produce more or less helpful measures but cannot be expected to yield anything approaching

a true “double” or “triple” bottom line. A true “bottom line” involves aggregation and comparability of costs and benefits and provides incentives to perform. The multitude of social and environmental measurement schemes will by necessity remain a patchwork that can be thought of as describing the “product characteristics” of a company’s output. Accounting profit remains the only measure that effectively aggregates costs and benefits and provides incentives. Profit itself is not just a necessity for organizational survival. It measures whether organizations meet client needs. It is thus an important measure of social impact in its own right. This may be unsurprising, but it sets expectations straight compared with currently widespread unrealistic hopes for the measurement of social and environmental impact and redirects attention to paying attention to profitability as part of impact measurement.

This paper is a product of the Inclusive Business Department, International Finance Corporation. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at muklein@gmail.com.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Firms Doing Good: How do we know?

Measurement of social and environmental results

Michael Klein¹ and Sumeet Kaur²

Keywords: Social Impact Investing, Corporate Social Responsibility, Results Measurement
JEL classification: D64, L21, M21

¹ *Johns Hopkins School of Advanced International Studies*

² *International Finance Corporation*

Contents

- I. Introduction..... 3
- II. Approaches to measurement..... 4
- III. Assessing Readiness 5
 - Organizational capacity*..... 5
 - Money spent for purpose intended*..... 6
- IV. Measurements of results 8
 - Reach* 8
 - Before and after*..... 8
- V. Making a difference: Measurement of causality and attribution 9
- VI. Finding the best option: Comparing organizations and tracking progress over time 11
 - Cost-effectiveness approaches* 11
 - Success rates and scoring approaches*..... 13
 - Cost-benefit analysis* 14
- VII. From measures to incentives – how many bottom lines? 15

- Annex 1 18
- Annex 2 19
- Annex 3 20
- Annex 4: The extent of social impact measurement 21
- Annex 5: Examples of social investment indices:..... 22
- References 23

I. Introduction

The modern firm emerged in the late 18th century. Ever since, efforts have been made to reconcile the productive power of capitalist firms with social and environmental concerns. Philanthropy, corporate social responsibility, impact investing and shared value investing are all attempts to focus firms in varying ways and to varying degrees on “doing good.” A key challenge for any such effort is to demonstrate that good is indeed done.

In recent years more and more measurement initiatives have sprung up that focus on assessing the good firms do.³ In 1997 the Global Reporting Initiative (GRI) was launched. In 2008 the Global Impact Investment Network (GIIN) was established and started developing Impact Reporting and Investment Standards (IRIS) as well as providing suggestions for indicators of social and environmental success.

Several organizations have experimented with improved methods of social and environmental impact measurement. The Acumen Fund developed a “Best Available Charitable Option” to measure its impact investing. The Roberts Enterprise Development Fund, a venture philanthropy organization, developed a measure of “Social Return on Investment.” The holy grail of measurement is sometimes called the “triple bottom line,” a term coined by John Elkington in 1995. As reflected in the Elkington inspired title of Shell’s first sustainability report in 1997, the goal is to measure impacts on “people, planet, profit.” Advocates imply that firms are to be guided by this triple bottom line as they were by profit so far.

The fundamental measurement challenges can be illustrated as follows. Imagine your child is sick and is hospitalized. Would it comfort you to know that the hospital is using the money it receives for buying the relevant medicine? Well, it is better to know that money is spent on the intended purpose than to find out that money is diverted for other things. Would it help to know that the medicine is actually given to the child? Yes, reaching the targeted beneficiary is a condition for success, but no more. Would it be good to know whether the child is better off after treatment? Clearly, but you cannot be sure that it was because of the treatment. Would it be good to know that the treatment was effective, because other children similar to yours that did not receive it fared worse? That is really what you want to know when you agree to an expensive treatment. If you now want to donate money to hospitals out of gratitude for healing your child, you would like to know not only that treatment is effective, but also that your money goes to the hospital that has the greatest success per dollar spent. For that you need to be able to compare hospitals’ costs. Or maybe you want to give money for the kind of treatments that help people most. Then you need to be able to compare not only costs, but also benefits from treatment of different illnesses. That means valuing the benefits and aggregating them to determine the “value proposition” of the hospital.

In practice it is rare that organizations are able to provide more than measures that assure supporters that money is spent on purposes intended or maybe that the targeted beneficiaries have been reached, i.e. only to the level of the first two questions in the above example. In an open letter to impact investors in December 2010 the GIIN spelled out a grand ambition for evaluation. “...impact investors must measure and report their social impact using rigorous standards that facilitate comparison and benchmarking. Social impact measurement must also draw attention to potential unintended

³ Methods to assess the social costs and benefits of projects go back to the work of Jules Dupuit in 1848. Governments started using social cost benefit analysis after World War II. International agencies, like the OECD and the World Bank, developed the method further since the late 1960s.

consequences and assess the full range of impact an investment can have – on investee businesses, customers, supply chains and the broader community.”⁴ In the following we review how far measurement has come to meet such ambitions.

II. Approaches to measurement

Measurement of social or environmental impact is most relevant for organizations that receive donations in some form. This includes customers who pay extra for the social or environmental value of a product such as fair trade coffee, investors who provide capital at below market rates to impact investing firms or “plain-vanilla” donors. It may also be relevant for firms that want to avoid reputational damage.

A whole array of methods exists to assess social and environmental impacts. Some are qualitative; others use or emphasize quantitative measures. Most actual evaluations mix qualitative and quantitative elements, depending on the problem at hand. Logical frameworks (“Logframes”) are often used to fit disparate pieces of information into reasonably coherent frameworks that help trace the effects of an action by a firm on stated goals.

Data are elicited from affected people through consultations, participant observation or interviews, others through measurement of physical, chemical or biological phenomena, for example to gauge the quality of the environment. Technology aids in the development of innovative measurement tools, for example, using online social networks, mobile phones, or new maps of social or environmental phenomena.

This paper provides a classification scheme which helps assess the various approaches to social impact measurement. The following basic measurement approaches for organizations pursuing social impact are distinguished:

- Measures of organizational readiness that shed light on an organization’s internal policies, organization and processes, as well as sources and uses of funds
- Measures of some form of result (“reach” and “before and after”)
- Measures that allow attribution of impact to the organization (“impact evaluation”)
- Measures that allow aggregation of results across projects, programs or organizations and thus also comparability between them (“cost-effectiveness” and “cost-benefit analysis”)

Finally incentive schemes may be tied to measures of results so as to provide or validate motivation to pursue social impact.

Any particular organization may, of course, mix elements of these metrics.

⁴ Global Impact Investing Network (2010) “Open letter from the Global Impact Investing Network”, December 9

III. Assessing readiness

Organizational capacity

Some measurements try to capture an organization's readiness to do good or the "character" of a firm. This may include information about firms' governance systems and ethical principles as well as organizational arrangements and procedures to implement them, including systems to cope with fraud or corruption. It may also cover a firm's labor and environmental practices.

Such measures dominate, for example, the family of Dow-Jones Sustainability Indices and the reporting requirements of the Global Reporting Initiative (GRI). The Dow-Jones Sustainability indices are long on characteristics of internal readiness assessment, for example, corporate governance arrangements, reporting practices, customer relationship and brand management, and human resource policies and processes.⁵

Most information requirements under GRI cover organizational arrangements and management approaches to economic, social and environmental issues. They include firm-internal arrangements to achieve good results, for example, procedures for hiring local staff, strategies to manage bio-diversity or employee training on human rights issues. Indicators on actual results achieved are sparse. The most concrete and well-defined measures are drawn from company accounts.

GRI also provides a rating method. This assesses the responsiveness to GRI reporting requirements based on self-assessment against the GRI criteria. It says little about whether the organization is correctly reporting or whether it is actually doing good (Box 1).

Box 1: GRI ratings

GRI is a leader in sustainability reporting because of its comprehensive guidelines on measuring economic, social, environmental, and governance indicators. However it is important to distinguish what GRI sustainability reporting is and is not. GRI's mission is to encourage companies to measure and disclose sustainability performance. The ratings of A, B, or C are awarded based on *level of disclosure* not actual *sustainability performance*. The data is largely self-reported; companies can elect to get "external assurance" about what they report and receive a '+' next to their rating, although it should be noted that "GRI...offers no opinion on whether the '+' is justified." These features allow for some surprising sustainability ratings, e.g. large oil companies such as Shell or Petróleos Mexicanos have ratings of 'A+' while a human rights organization such as Amnesty International (Secretariat) has a rating of 'C'. Without a deeper analysis, the first reaction to a rating system is that a higher rating is related to higher social and environmental impact, but the reality could be different and more reflective of the amount of resources a company has to dedicate to reporting mechanisms.

⁵ <http://www.sustainability-indexes.com/dow-jones-sustainability-indices/index.jsp>

The focus on organizational readiness makes it possible to apply reporting principles to a large array of disparate organizations operating in any sector of an economy. It is also hoped that firms that have policies as well as organizational groups and processes to implement them are actually likelier to achieve good outcomes than firms that do not. To assess whether this is indeed true would require precisely the outcome measurement that is typically not much covered by readiness assessments.

The Dow-Jones Indices allow a check whether organizations that are “ready” are actually performing well financially by studying the financial performance of “sustainably managed” firms. Even there it is hard to figure out what exactly drove performance. For example, corporate governance research has so far not been able to identify clear causal linkages between economic performance and approaches that are deemed to be best practice, such as the number of independent directors on boards or the separation of CEO and chairman.

The links between readiness measures and social and environmental outcomes are even harder to assess. Examples like that of Enron, which had “best practice” corporate governance and ethical policies, provide a warning that the trappings of readiness may not translate into good outcomes. The Dow-Jones Sustainability Indices take this into account through a procedure that questions a company’s rating when media reports of allegations and scandals arise.

Overall readiness assessments measure the “athlete’s muscle tone but not how fast she can actually run the 100 meters.”⁶ That is potentially valuable information, but not a clear indication of whether a firm actually achieves good outcomes. In fact it would make sense to clearly delineate readiness measures from outcome measures so that one can assess whether and how readiness translates into results.

Money spent for purpose intended

Related to readiness assessments are ways of determining whether an organization actually spends money on the stated purposes. Relevant metrics help assess whether there may be fraud or waste in an organization.

Sometimes organizations simply do not know whether or how much money has been spent on stated purposes. For example, some microfinance organizations assessed by the micro-finance donor group, CGAP (Consultative Group to Assist the Poor), were not able to show what they spent on microfinance activities.⁷

In other cases data cast doubt on organizational effectiveness and possibly on fraud. For example, a recent scandal in the US revealed that InfoCision, a telemarketer, raised money for various not-for-profit firms. In the case of the American Diabetes Organization (ADA) the telemarketer retained 85 percent of donations under its contract even though the “sales pitch” claimed that on average 75 percent of all funds received by ADA went directly to beneficiaries of the charity.⁸

⁶ In the words of former World Bank President Paul Wolfowitz.

⁷ Personal communication to the author. CGAP (2006) presents a subset of findings from a broader review of several aid agencies, which contained the quoted finding.

⁸ Bloomberg News (2012)

Basic accounting data including sources and uses of funds can thus help provide valuable information for donors of one type or another. First, donors can learn whether relevant data exist. Second, they can learn what share of expenses went on clearly identifiable expenses such as fundraising. At the same time the “share of administrative expenses” that some organizations provide can be highly misleading and is not comparable across organizations. Classifying expenses as administrative provides ample scope for judgment calls. Related to this, some organizations such as advocacy NGOs, spent almost all their funds on some form of administrative expenses, because that is the nature of advocacy work. Others mainly pass money on to beneficiaries, like some disaster relief organizations. Nevertheless providing such cost data can help donors form their own judgment on whether they like what an organization is doing with their money.

Accounting data may, of course, simply be misreported. For this reason accounting standards establish credibility and help to spell out how numbers were derived, and auditors verify those numbers as reputable third parties. At the same time accounting and auditing scandals in the commercial world attest to the limits of any measurement system. After all financial accounting and auditing is a highly developed discipline compared to the nascent field of social and environmental results measurement.

Box 2: Charity Navigator

Charity Navigator is a charity rating agency that provides an independent evaluation of a charity’s effectiveness. They use two criteria for their measure: the financial health of a charity and its accountability and transparency. Using a charity’s publically available tax returns (IRS form 990) and information posted on its website, Charity Navigator provides each charity a rating ranging from zero to four stars. Charity Navigator provides this rating for 5,400 U.S. charities; they want donors to be able to compare across charities on questions such as administrative expenses, fundraising expenses, good governance practices, and transparency of information.

This method gives a high rating to a charity with lower overhead costs, low CEO compensation, and greater transparency, which may appeal to a donor whose intent it to channel most of the funding directly to the intended recipient. However there are some important unresolved issues with this approach. One is that the IRS form 990 does not have strict reporting standards so it is easy for charity organization to under- or misreport their fundraising expenses. A second issue is that some charities argue that CEO and staff salaries have been to be competitive to attract top talent, otherwise an organization may appear to be administratively efficient in costs but is not efficient in delivering impact. This relates to a third shortfall of this approach which is that it is focused solely on internal capacity, and does not address the outcomes or impact of a charity’s work. In January 2013, Charity Navigator introduced a Results Reporting assessment, which is still being formulated and is expected to be launched in 2016.

IV. Measurements of results

Reach

A large part of the measures collected, for example by IRIS, are about the reach of an organization. These measures cover things like the number of customers served, units sold or installed, deals closed, people trained, training books sold, and a myriad of similar indicators.

Reach numbers are very popular. For example, microfinance organizations like reporting the number of clients reached. This constitutes real progress as a number of organizations are not capable of reporting this. At the same time such measures do not say anything about whether an organization was actually effective. Reach numbers simply say that the organization focused on desirable purposes.

Before and after

One step up from reach indicators are those capturing program effects or outcomes before and after an intervention, for example whether recipients of microfinance actually did better on some measure after the intervention. Change in reach indicators provides the simplest “before and after” measure. If one simply measures a change after an intervention, then we learn that an effect was possible, but have no way of telling whether there actually was any.

Relevant indicators as suggested, for example by IRIS, might cover things like trees planted, energy savings achieved or jobs created. Outcomes such as these are intermediate outcomes. For example, trees might be planted to enhance the absorption of greenhouse gas emissions. Such outcomes further down the line are often called impacts. In practice most forms of outcomes, from proximate to final ones, are various forms of intermediate outcomes – somewhere along the continuum to world happiness.

Proximate or intermediate outcomes are often more easily measured than more remote or final outcomes. Also proximate outcomes may more easily be attributed to the intervention by a particular organization. While some claim impact or final outcome measurement should be the goal, in practice this can be a case of the better driving out the good.

The issue is what can sensibly be measured and acted upon. At the same time it is necessary to have at least a notion how an intermediate outcome matters ultimately. This may be based on theory and anecdotal knowledge, and the occasional in depth impact evaluation study that attempts to link outcome measures to long-term impact. As we learn more about how the world works such “notions” will need to be updated. If one could fully trace the effect of any firm’s intervention all the way to final impacts, one could just as well advocate central planning, because all the noxious information problems that plague us in reality would have been solved.

Reach numbers may be of dubious quality. For example, many organizations would like to argue that they helped create new jobs. Yet such numbers are hard to come by. Many firms employ people under a variety of contractual forms. Some are full-time, some part-time, some consultants, some interns, etc. Some contracts are purposefully designed to work around formal limits like complement controls or

labor laws. Almost by definition it is hard to generate numbers on such employment contracts, not to speak of calculating full-time equivalents.

One measure suggested by IRIS is “jobs maintained” expressed in full-time equivalents. This captures official employment at firms, even if the measurement of full-time equivalent may be highly dubious in many cases. It does not imply that firms added these jobs during a particular year or that the organization providing support was the cause of it. The measurement thus simply establishes that an organization might have helped with job creation.

Reach numbers can sometimes be combined, for example, with accounting data to produce simple indicators of performance. Examples are number of clients reached per sales representative or the average cost per job maintained. This allows rudimentary assessments of effectiveness. For example, New Profit, a venture philanthropy fund, assesses its portfolio investments using two measures of compound annual growth rate for “revenues” and “lives touched” – “revenues” being a proxy for performance and “lives touched” being a proxy for social impact. In 2010 New Profit reports a CAGR of 41% and 40% respectively for each of these measures. Like other reach measures, this is a step in the right direction, but should not be mistaken for indicating actual social impact since investors have no mechanism for assessing whether people’s lives actually improved through these programs.

Some venture philanthropy funds employ “before and after” measurements. The Robin Hood Foundation, which invests in education and job training programs in New York, employs a ‘benefits-cost ratio’ that measures the increased salaries of those enrolled in the grantee programs. Ashoka administers a ‘measuring effectiveness questionnaire’ to the social entrepreneur fellows supported through its funding; the questionnaire uses four proxies of success including whether there has been a policy change in the institutional environment as a result of the fellow’s contributions to the community. The Grameen Foundation’s ‘Progress out of Poverty Index’ measures changes in indicators of poverty of households after receiving a microfinance loan, e.g. changes in the number of durable goods in a household, fuel type, or flooring type in the house.

“Before and after” measurement can tell us something about whether matters improved or deteriorated. Yet, we cannot be sure that it was the intervention that made the difference. The observed outcome might just have happened in any case.

V. Making a difference: Measurement of causality and attribution

The next level up in measurement is able to capture whether a project, program or organization is making a difference. For this we need to know what would have happened in the absence of an intervention, i.e. the counterfactual. For example, patients may have recovered after medical treatment, but it might have been a natural recovery that would have happened anyway. By comparing the treated patients to a control group of similar patients who were not treated we can see whether an intervention made a difference.

We thus need to measure not only what happens to the beneficiaries of an intervention but also what happens to others that form the control group. This increases measurement effort and costs. It is often neglected in the course of ordinary business of most well-meaning organizations; those that are aware

can rarely afford to carry out such evaluations. In fact, most do not routinely establish baseline data for the treatment group, much less for a control group as well.

In addition the composition of the control group matters. Consider, for example, small- and medium-scale enterprises (SMEs) that are meant to benefit from an intervention. If the treatment group does better than the control group after an intervention, it might not be due to the intervention. It could also be that the good outcome reflects the fact that good managers of SMEs were clever enough to avail themselves of the intervention, whereas others were not clever and did not figure out that help was available. The real driver of success might then have been the quality of management and not the treatment.

Conducting even simple forms of randomized controlled trials (RCT) or so-called quasi-experimental trials⁹ can make a major difference to evaluations. The International Finance Corporation at one time provided advice to seaweed farmers in Indonesia. Evaluations showed that the advice was competent and that farmers were better off afterwards. Later a quasi-experimental evaluation was conducted evaluating results against a control group. It turned out other farmers who had not received the advice had learned as well and performed just as well or better. They probably benefitted from advice given by agents of firms purchasing seaweed. It may thus have been a waste of money to fund advice; the market might have solved the problem in the course of normal operations.¹⁰ Alternatively, the good advice may eventually have reached farmers who did not receive it directly by word of mouth, although that would not explain outperformance by that group.

The gold standard of impact assessment, therefore, is the use of control groups, ideally randomized controlled trials (RCT). These are by now fairly routine for lots of pharmaceuticals, where we really want to know whether a medicine works or not. Note, however, that it took a long time for such methods to be adopted for medicine. Already at the turn of the 16th century Jan Baptist van Helmont (1579-1644) proposed RCTs: “Let us take out of the Hospitals, out of the Camps, or from elsewhere, 200, or 500 poor People, that have Fevers, Pleurisies, etc. Let us divide them in Halfes, let us cast lots, that one half of them may fall to my share and the other to yours; I will cure them without bloodletting and sensible evacuation; but do you do as ye know...we shall see how many Funerals both of us shall have: But let the reward of the contention or wager, be 300 Florens, deposited on both sides: Here your business is decided.”¹¹

It took several hundred years for such trials to become commonplace for testing the effectiveness of medicine, although not for testing that of doctors or hospitals. Today the methodology is better established. In recent years a number of RCTs have been undertaken, for example, in the field of health, education and microfinance, promising to shed light on what really works in these areas.

⁹ For example, in the absence of randomly selected control and treatment groups, one can try to capture the essential attributes of participants in the two groups and adjusted statistically for observable differences through methods such as propensity scoring. This, of course, still leaves open whether unobservable differences might account for the observed effect of an intervention rather than the intervention itself.

¹⁰ In this case this is also a way of assessing what some official development organizations call “additionality” namely whether the market or other organizations would do anyway what they set out to do. Taxpayer money should after all be spent on interventions that would not otherwise happen.

¹¹ <http://ije.oxfordjournals.org/content/30/5/1156.full>

While RCTs can make a convincing case for the effectiveness of certain interventions, unsurprisingly they still have their limits. It could, for example, be that both treatment and control groups have common characteristics that limit the transferability of lessons to other settings, for example other countries (this is called “external validity”). They may thus not be able to overcome some of the short-comings of standard statistical tests such as cross-country regressions. It may be impossible to truly separate control and treatment groups. In fact, the whole point of a business intervention may be to pick the best clients and not a random group.

Most RCTs tell us little about the costs of interventions and whether the costs were worth the benefits. For many important matters RCTs can simply not be used. For example, at the extreme RCTs cannot be conducted for a global carbon emissions tax. One would have to select randomly different, but similar planets and accept that some of them may be doomed to undergo global warming. Critical issues for RCTs here are whether systemic interventions can be subjected to them and whether the ethics of doing so are acceptable.¹²

VI. Finding the best option: Comparing organizations and tracking progress over time

Once we know that an intervention works, we may still want to know what it costs and whether it is worth the benefits. Only when we know the cost-benefit calculus can we quantitatively compare projects, and only when we can compare and aggregate interventions can we choose among programs or organizations. Comparability across time is needed to track progress. Finally, comparing costs and benefits allows us to assess whether financial profit is the right measure of success or whether there is a trade-off with social or environmental concerns and what that trade-off looks like. Thus to meet the spirit of the GIIN letter quoted at the beginning we need to go beyond impact assessment and also capture costs and benefits in comparable ways.

Cost-effectiveness approaches

In the face of the formidable problems to meet these demanding measurement challenges, most firms, sensibly, try to make progress with partial measurement approaches. Some try to use cost-effectiveness measures, such as the Center for High Impact Philanthropy’s “cost per Impact” method.

The Acumen Fund has developed a measure of the “Best available Charitable Option” (BACO). This metric benchmarks (quasi-)commercial ventures against pure grant funded ones. Take the example of bed nets for protection against malaria. At one end of the spectrum a pure grant funded organization may be able to supply a certain number of poor people with bed nets. When an organization is able to charge poor people to some extent for bed nets a given amount of donations can help reach a larger number of poor people.¹³

The best results under the BACO metric would be obtained by an organization that has been able to develop a commercially viable business without any need for donations. A commercial supplier of goods for low income people, like Walmart, would tend to look good on the metric. When actions are guided

¹² See Ravallion (2012) for a discussion of the limits of RCTs.

¹³ Sale of items also provides greater certainty that beneficiaries are reached who value the item and it provides an audit trail for results measurement.

by the metric they may also lead organizations to develop a business for those poor people who have the highest purchasing power and neglect others.

The metric is useful to provide a sense of which organization is closest to develop a sustainable business at the “bottom of the pyramid.” It is related to a method to calculate “subsidy dependence” that has been used in assessing to what degree credit schemes for agriculture, small and medium enterprises or microfinance are commercially sustainable (Schreiner M. and J. Yaron, 1999).

The Roberts Enterprise Development Fund (REDF) has developed a measure of social return on investment (SROI) that adjusts financial results for social benefits not captured by market prices. REDF funds organizations that help unemployed or homeless people gain employment. Essentially, SROI adjusts the financial results of the supported organizations by the saving in public sector resources that the government would otherwise be providing as social security to the unemployed or homeless. Such public savings are counted as an increase in social revenues and improve the return on investment measure.

Another way of thinking about the problem draws on standard “shadow” pricing methods from cost-benefit analysis. REDF-supported organizations demonstrate that previously unemployed people are able to produce work generating revenue. However, their productivity is lower than the cost of their wages. Hence a need remains for some subsidy. The shadow wage is thus below the wage they are being paid. If we were to compare the public welfare system to the REDF approach we would find that the public system has a lower social return on investment than REDF because it requires more subsidies (tax-funded in this case). The SROI would, however, look even better for firms employing the previously unemployed, if labor markets were more flexible and wages were set at a level corresponding to productivity. In that case firms would be willing to employ people without any subsidy. Even if one thought market wages were below a decent standard of a “living wage” and if one wanted to provide a subsidy to bring the total income of workers to that level, a system with flexible labor markets and subsidies in the form of a negative income tax might well yield better results and show. The SROI measures would confirm that.

Like BACO the SROI is a measure that may lead organizations to help reduce dependence on grants of some type. Again an organization such as Walmart would tend to look good on such indicators as it employs many people who might otherwise be unemployed. Implicit in the above measures is the view that pure market based, unsubsidized solutions would be best. Both measures would tend to favor focusing help on the best-off of the poor, because they might be more likely to purchase services and gain employment.

Another approach has been developed by the University of Pennsylvania’s Center for High Impact Philanthropy, which uses a ‘Cost per Impact’ measure to determine social impact. The impact of the project is based on matching outputs and outcomes with previously conducted RTCs, quasi-experimental research, and other published studies. This is an example of a complex approach mixing benchmarks obtained from prior studies with actual program information. Whether past studies are fully comparable to the programs being evaluated remains open.

The measures that have been tried can only incompletely capture all costs and benefits. Measurement problems abound. For example, how do we measure whether a person employed under an REDF program would really be unemployed otherwise? The metrics may also produce certain biases that one would need to adjust for when making decisions. Guidance for decision-making is thus limited. Given

the cost of calculating the indicators, it is not too surprising, and probably sensible, that the pioneering metrics are not calculated and published consistently.¹⁴

Success rates and scoring approaches

Various approaches to aggregating information on results have been tried. Several development organizations and foundations¹⁵ report “success rates” for their activities. This requires a system of assessing the results of activities through a mix of quantitative and qualitative measures, weighing them and rating them as successful or unsuccessful. The percentage of successful activities is then reported as the success rate.

In interpreting success rates one needs to be aware of how they have been derived. For example, some organizations may not evaluate a random sample of activities but disproportionately those that appear successful. Even when evaluations are done for a random sample of activities there may be systematic biases across organizations. For example, in the field of official development organizations some tend to report success rates around 80 percent. That includes the World Bank and in Germany, GIZ. At the same time the private sector arm of the World Bank, the International Finance Corporation, reports much lower success rates, around 65 percent. The difference between KfW, the German development bank and GIZ, the official German development technical assistance provider is of a similar nature.¹⁶

Does this mean that GIZ is better than KfW? When looking at the details the following picture emerges. Organizations that have a substantial portfolio of (quasi-)commercial operations that operate in markets tend to show success rates of around 50 percent or less for the financial success of such operations, which typically forms part of the basis for evaluating success. Success with regards to financial performance often means that the organization’s financial results have surpassed the cost of capital in the market. The cost of capital is de facto a measure of the average performance of firms in the market. An organization that systematically beats the market more than 50 percent of the time is, indeed, rather successful. One way of enhancing results is thus to choose a very low cost of capital for evaluations as some development organizations have done.

Organizations like the World Bank and GIZ, on the other hand, mostly pursue activities that do not financially compete in markets. The typical, and often sensible, approach to evaluation is then to compare results to stated goals. When goals are modest, results are good. Also, there is substantial scope for discretion in judging results. A metric like the cost of capital that compares the results to a form of control group usually does not exist. Hence we tend to find high success rates. Facetiously put, imagine one were to evaluate the performance of bakeries under central planning compared to market economies. One would find high success rates under central planning, because most bakeries would make edible bread. In market economies, however, most bakeries would underperform the market and less than 50 percent would achieve above average results.

Another approach comparing organizations’ overall results is to use more complex scoring methods. These would aggregate results for different parts of the portfolio of activities of an organization by weighing them in some fashion and then adding them up. This is the approach pursued by several

¹⁴ Tuan (2008) provides examples of intermittent use.

¹⁵ E.g. many aid organizations including the World Bank or some NGOs like the Amber Foundation: www.amberweb.org/

¹⁶ Michaelowa and Borrmann (2004)

development organizations in Europe, of the so-called EDFI (European Development Finance Institutions) group. However, the chosen weights are by necessity subjective and opinions about the relative importance of different activities obviously vary. In the case of the EDFIs, different organizations also pursue different goals with different weights. It is thus not really possible to compare these organizations just on the basis of the overall score.

As weights are typically arbitrary they may lead to counter-intuitive results. For example, the European Organizations found a few years ago that their weighing system discriminated against infrastructure. They thus changed the weights. This renders tracking of progress over time quite difficult.

Cost-benefit analysis

The conceptually most advanced method to compare projects and aggregate results is cost-benefit analysis. Development aid organizations such as the World Bank Group have developed practices and use the method to varying degrees.

Calculating costs and benefits requires a way to weigh the different elements of costs and benefits, using what economists call a “numeraire.” In functioning markets, prices provide such a numeraire. In fact, markets solve a great deal of the measurement challenges discussed so far. Customers decide whether the goods or services provided by a firm are likely to make them better off and whether they are worth the costs. Even where the supplying firm is a monopolist, as long as customers are not forced to buy they reveal that they are better off by purchasing. When customers freely choose among competing suppliers the market also establishes which firms are most efficient in supplying what customers want. Competitive markets thus solve all problems discussed so far as long as the purchasing decisions of buyers reflect all the relevant benefits of what firms offer. In the words of the Sociologist Georg Simmel: “Modern competition is described as the fight of all against all, but at the same time it is the fight for all.....Innumerable times [competition] achieves what usually only love can do: The divination of the innermost wishes of the other, even before he himself becomes aware of it”.¹⁷

Profit or rather the value of the firm – expressed in share prices when firms are tradable – is then a measure of whether the firm takes care of customers’ needs. Because the providers of goods and services cannot live without fulfilling customers’ needs, competitive markets provide not only a measurement system for meeting peoples’ needs but also an incentive for firms to care about these needs.

In debates about some form of “social business” one often hears the argument that firms have to make profit to be able to sustain operations that allow the firm to do good. Profit is thus often cast as a necessary evil to enable good things to happen. The financials have to work for the social and environmental impact to be achieved. One can call this the “prostitution theory of profit.” To do good one has to do questionable things like pursuing financial gain. In fact, profit measures whether needs have been met. Even when goods and services are sold by monopolists, profitability tells us that customers preferred having the service over going without it. That is a critical measure of social impact.

Profit fails as a measure of meeting social and environmental concerns when costs and benefits of a firm’s activities are not priced and included in the profit calculus. Economists call this “externalities.”

¹⁷ Quoted in Hirschman A. (1982)

For example, when a firm pollutes the environment, but the costs of such pollution are not priced and charged to the firm, profit provides neither a measure of impact nor an incentive to behave well. On the contrary if goods can be made cheaper or better by polluting, more firms have an incentive to damage the environment.

It may also happen that customers do not act in their own best interest. Most clearly that can happen, when children are very young or when grow-ups are mentally impaired in some form. One can also argue that addicts belong into this category. In this case customer demand expressed in the market may not reflect social value from a broader perspective. People may also suffer from information problems and biases, which may render market prices less meaningful.

For decades economists have tried to find ways of pricing externalities and paternalistic preferences. This gave rise to social cost benefit analysis. Good and bad things that are not captured by market prices are valued with so-called shadow prices. A new bottom line is then derived that measures profit adjusted for these costs and benefits. For example, to capture the impact of greenhouse gas emissions one may estimate the environmental cost of emitting a ton of carbon and treat that as a cost. Or one may assess the time-saving benefits from a road project by valuing the time gained. One can also assess who gained and who lost from an activity and weigh such effects based on social preferences about poverty or inequality.

Such social cost-benefit analysis assesses the impact of activities on all stakeholders – customers, workers, other affected parties and the environment. It is conceptually the answer to the measurement challenge posed, for example, by GIIN. It may benefit in turn from control group based evidence about impacts, which can be subjected to valuation.

Measuring costs and benefits without actual market prices is, however, fraught with uncertainty and arbitrariness. For example, estimates for shadow prices for greenhouse gas emissions can easily vary between, say, 20\$/ton of carbon and a 100\$/ton depending on the abatement cost curves one assumes. Alternatively benefits may be assessed by contingent valuation methods. Here possible interested parties are asked what they would be willing to pay if a new service were to be offered or what taxes they might be willing to pay if certain public goods were to be provided by government. Incentives to misrepresent abound and people may just make up answers on the fly. While such valuation may be better than simple guesses, results from such stated preference methods are fraught with substantial error margins and can at times also be close to meaningless (Journal of Economic Perspectives 2012).

If social cost benefit analysis could be perfectly deployed it would allow funders to compare different organizations aiming at social or environmental benefit more easily, just as financial results allow analysts to compare the performance of commercial firms. However, the generic problems of measuring impact and valuing non-market costs and benefits severely limit this possibility. Just like REDF stopped calculating SROI, many development and government organizations reduced the emphasis on full-fledged cost-benefit analysis (Little and Mirrlees 1990; Hahn and Dudley (2007).

VII. From measures to incentives – how many bottom lines?

Financial results, profits or losses, are commonly called the bottom line for a firm. The term “bottom line” implies that all costs and benefits are aggregated and summarized in a single metric “profit” or

better still the value of the firm that reflects expectations about future profitability.¹⁸ Importantly, it also means that the summary measure drives behavior of the firm.¹⁹

Even where social cost-benefit analysis is feasible, it still does not provide the same incentive to firms as financial results do. It may, for example, be the case that firms lose money if they follow the direction indicated by social cost benefit analysis. In that case they have no incentive to heed its guidance. Only when firms are financially compensated for the departure of social costs and benefits from financial costs and benefits, would there be an incentive to do as SCBA says. This might happen in theory if some stakeholders were willing to compensate the firm voluntarily, if government provided an equivalent mix of taxes and/or subsidies, or if actual markets were created to value all relevant external effects.

In practice the few existing evaluations about voluntary compensation of firms indicate that people are not willing to pay much to compensate firms for costs and benefits to stakeholders that are not priced in markets. Firms that try to establish a reputation for treating customers well are rewarded with better prices. Firms that try to protect the environment are typically not receiving significant voluntary payments from stakeholders, such as customers, and lose money unless environmental stewardship also translates into lower costs, for example, through energy efficiency (Mayer 2013).²⁰

After reviewing all basic approaches to results measurement, we are left with the conclusion that there is no way around a significant amount of subjective judgment when evaluating organizations, comparing them and tracking their progress over time. Individual indicators and measurements can help inform such judgment. However, no measurement approach seems in reach that could rival the power of financial accounting when it comes to aggregating the results from activities, benchmarking them against other organizations and providing incentives to act accordingly. This may not be surprising, but it serves as a note of caution when setting goals for results measurement such as the ones expressed in the GIIN letter quoted in the beginning.

At the same time there is no need to despair. Consider a normal commercial market, for example, for mobile phones. The performance of mobile phones varies on many dimensions. Companies provide information about technical characteristics, about compatibility with other devices, about the carriers offering service and so on. They also market stories about the product to make customers feel good about it. Customers may be helped by rating services such as consumer reports. Customers look at all this to varying degrees. They then make their own judgment whether the product is worth the cost. They thus value the product by mixing heterogeneous information in the way they want.

It is no different for organizations aiming at social or environmental impact. In some cases they may be commercially fully viable. They are then in the same situation as other companies, and after all companies can only make profit if they satisfy needs. Other organizations may require some form of subsidy. They may need consumers willing to pay extra for “fair trade” coffee. They may require

¹⁸ see for example Berk and De Marzo (2014)

¹⁹ Norman and McDonald (2003) do not cover the incentive compatibility aspect

²⁰ What does provide an incentive to pursue social or environmental goals are appropriate taxes or regulations backed by adequate penalties. Alternatively, it may in some cases be possible to establish properties rights, for example, for SO₂ emissions that lead to the creation of a market price for emitting such pollutants. All this requires government intervention.

investors who are willing to accept below market returns, or they may just try to attract straight donations. These may all be thought of as donations of some form.

Those providing donations are like customers. They buy a service that benefits someone other than themselves. They would like to know if buying it is worth it. Therefore organizations provide information about the social and environmental characteristics of their operations or the goods and services they provide. There will be a mix of quantitative information, argument about the quality of the organization, stories of success and the like.²¹ Customers then make their purchase (donation) decision. Successful companies will see that success reflected in their financial bottom line.

The financial bottom line is the only one that can usefully be called “bottom line” with all its connotations about aggregation and incentives. The talk about “double” or “triple” bottom line may be cute, but is hardly enlightening. Social and environmental information often ends up as a marketing act, no different than any other firm trying to make its case. It is thus important for organizations that want to attract donations to make their case and measure results in a sensible way. However, nobody should be under the illusion that anything like a financial bottom line will be the result of such efforts.²²

²¹ Ebrahim and Kasturi Rangan (2010) also make the point that pragmatic and selective measurement approaches are the way to go.

²² See also Norman and MacDonald (2003) and Altman and Berman (2011) for views that multiple bottom line concepts are conceptually meaningless.

Annex 1

Global Reporting Initiative (GRI):

About: “The Global Reporting Initiative (GRI) is a non-profit organization that promotes economic, environmental and social sustainability. GRI provides all companies and organizations with a comprehensive sustainability reporting framework that is widely used around the world.”

Vision: “A sustainable global economy where organizations manage their economic, environmental, social and governance performance and impacts responsibly and report transparently.”

Mission: “To make sustainability reporting standard practice by providing guidance and support to organizations.”

GRI’s Framework consists of the Sustainability Reporting Guidelines, 12 Sector Supplements, National Annexes (only Brazil pilot thus far), and the Boundary and Technical Protocols

GRI 3.1 has three “standard disclosures”:

- I. Profile Disclosures
 1. Strategy and Analysis – 2 disclosures
 2. Organizational Profile – 10 disclosures
 3. Report Parameters – 13 disclosures
 4. Governance, Commitments, and Engagement – 17 disclosures
- II. Disclosures on Management Approach – 37 disclosures
- III. Performance Indicators – total 81 indicators – 11% economic, 37% environmental, 52% social
 1. Economic – 9 indicators
 2. Environmental – 30 indicators
 3. Social: Labor Practices and Decent Work – 14 indicators
 4. Social: Human Rights – 11 indicators
 5. Social: Society – 8 indicators
 6. Social Product Responsibility – 9 indicators

Some examples of sector supplements:

1. Financial Services: additional 16 indicators, and new performance indicators area called “Product and Service Impact.” Examples of indicators: “Monetary value of products and services designed to deliver a specific social or environmental benefit for each business line broken down by purpose,” “Access points in low-populated or economically disadvantaged areas by type,” or “Initiatives to enhance financial literacy by type of beneficiary”.
2. NGO: additional 9 indicators, and new performance indicator area called “Program Effectiveness.” Examples of indicators include resource allocation and ethical fundraising.
3. Oil & Gas: additional 14 indicators, no new performance indicator area. Examples of indicators include total amount invested in renewable energy and about involuntary resettlement.

Annex 2

Ashoka's Measuring Effectiveness Questionnaire:

Ashoka is a nonprofit organization based in Arlington, VA that supports social entrepreneur fellows. Currently it operates in over 70 countries and supports the work of 2,000 social entrepreneurs. Information about each fellow, the problem they were trying to address, and their strategy used is available publically on Ashoka's website, with records available from 1982-present.

On an annual basis Ashoka sends a Measuring Effectiveness Questionnaire to the class of Fellows elected five and ten years prior. The questionnaire is two pages long and has eight questions. Ashoka measures its success by the degree to which it has empowered its fellows to impact change in their environment. The questions ask whether the fellow, five or ten years after the support from Ashoka, is still working towards his or her original vision, whether others have replicated the idea, whether the fellow's work has changed public policy, and how the fellow perceives the support provided by Ashoka. After the questionnaire, some fellows are selected for interviews.

The questionnaire chooses proxy indicators that measure Ashoka's success via the ultimate outcomes desired, i.e. to what extent there has been positive change in the society where the fellow lives. The measurements are qualitative, self-reported by fellows, and unverified by a third party. Ashoka has aggregated the survey results from the past six years into a "Measuring Effectiveness" report, which shows very positive results, e.g. five years post election 94% of fellows are still working towards their original vision and 56% have impacted public policy; ten years post election the numbers are 83% and 71% respectively. It is unclear if these are the numbers from fellows who filled the questionnaire or all fellows supported through the program (the questionnaire response rate was 83% for fellows elected 5 years ago and 68% for fellows elected ten years ago). Ashoka's "Measuring Effectiveness" report does not provide information on the dollars invested, and so lacks an analysis of cost effectiveness for the results (which is useful for comparability across investments in individual fellows or across other organizations similar to Ashoka). However Ashoka's financial and audit statements are publically available on their website.

On the "Impact" page of the Ashoka website, the impact of four fellows is quantified, e.g. one fellow "helped cut rural electrification costs by 70-80% in Brazil and brought electricity to the homes of over 1 million people. Fabio's innovation has spread to 23 countries worldwide."

Annex 3

William and Flora Hewlett Foundation's Expected Return:

The William and Flora Hewlett Foundation is one of the largest grant organizations in the US. The foundation has assets of over \$7 billion and invests in education, global development, the environment, and local communities in San Francisco, among other areas.

The Hewlett Foundation employs an "expected return" methodology to evaluate potential grants, and invests in those that offer the highest expected returns. Expected return is calculated as the benefit multiplied by the likelihood of success, divided by the cost.

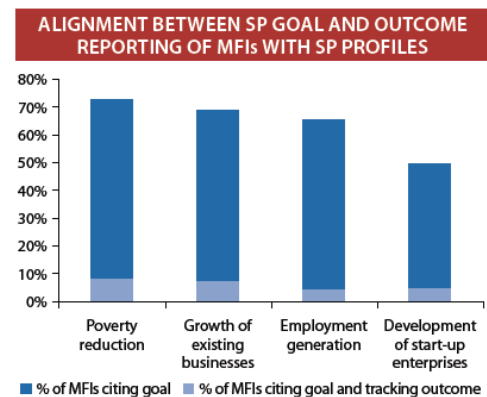
The methodology first defines the yardstick with which to measure the benefits. For example, in the field of global development this was accepted as the number of people living on less than \$2/day; this was further expanded into a more comprehensive index that included literary and health indicators. The benefits are estimated using existing academic research; for example a program that invested in education drew on research that estimates the impact of higher education levels and income. The benefits are then discounted with a likelihood of success. This is measured as a combination of strategic accuracy, grantee success, and external conditions, and essentially tries to make explicit the inherent risks of every program activity. The foundation estimates these based on interviews with estimates and based on past experiences. The benefits are discounted by the likelihood of success and then divided by the total costs such includes the program costs and overhead expenses, which yields an "expected return" for each program. The 2008 report describing the expected return methodology gives on example of a calculation whereby "every \$1 million that the Hewlett Foundation spent on improving technical assistance in Nigeria would likely contribute to doubling of income for about 6,700 poor people."

This methodology is useful for comparing the cost effectiveness of different programs and even across organizations, however with a recognition of the inherent difficulty in quantifying the benefits and risks of program activities. The Hewlett Foundation does not report on the expected returns of its programs. The website provides a listing of the dollars its has invested, and the annual reports provide descriptive measures of success, academic research showing improvements in outcome areas, and output numbers from foundation-supported programs such as the increased use of contraceptives in an area or the number of African masters and doctoral graduates.

Website: <http://www.hewlett.org>

Annex 4: The extent of social impact measurement

- There is no good estimation of the number of methodologies being employed. The Foundation Center’s Tools and Resources for Assessing Social Impact, which is quite comprehensive, contains a database of 190 tools, methods, and best practices. (Source: TRASI Database <http://trasi.foundationcenter.org/browse.php>)
- GRI’s mission is “to make sustainability reporting standard practice for all organizations.” The GRI Sustainability Disclosure Database contains 13,225 reports from 5,124 organizations. (Source: <http://database.globalreporting.org>).
- Ninety-five percent of the 250 largest companies in the world (G250 companies) now report on their corporate responsibility activities. 80% of reporting companies use GRI. (Source: “KPMG International Corporate Responsibility Reporting Survey 2011.” <http://www.kpmg.com/PT/pt/IssuesAndInsights/Documents/corporate-responsibility2011.pdf>)
- ISO 14001 (requirements for environmental management systems) issued 267,457 certifications in 2011. No data yet for ISO 26000 (social responsibility standards) because it was initiated in December 2010. (Source: “ISO Survey of Management Systems Standard Certifications – 2011.” http://www.iso.org/iso/iso_survey2011_executive-summary.pdf)
- Microfinance Information Exchange (MIX) had 2,100 MFIs from 110 countries reporting data in 2011-2012. (Source: “MIX Fiscal Year 2012 Annual Report.” http://www.mixmarket.org/sites/default/files/fy2012_mix_annual_report.pdf)
 - Chart from report based on 405 MFIs:



Annex 5: Examples of social investment indices:

- Calvert Social Index - Calvert's Sustainability Research Department analyzes the 1,000 largest companies in the U.S. A sustainability audit is conducted in the following areas: governance and ethics; environment; workplace; product safety and impact; community relations; international operations and human rights; and indigenous peoples' rights.
- FTSE4Good - Research for the indices is supported by the Ethical Investment Research Services (EIRIS).
- MSCI ESG Indices
- Domini 400 Social Index
- Dow Jones Sustainability Indices
- Environmental focus:
 - S&P Global Clean Energy Index
 - S&P Global Water Index
 - WilderHill Clean Energy Index
 - Cleantech Index

References

- Acumen Fund. "The Best Available Charitable Option." Website link: http://www.impact.upenn.edu/about/cost-and-impact/linking_considerations_of_cost_and_impact/
- Altman D. and Berman J. (2011) "The Single Bottom Line" Stern School of Business and Dalberg Global Advisors, mimeo, June 13
- Ashoka Innovators for the Public. (2006). "Measuring Effectiveness: A Six Year Summary of Methodology and Findings." Website link: https://www.ashoka.org/files/ME_Impact06.pdf
- Berk and De Marzo (2014) "Corporate Finance" 3rd edition, Pearson
- Bloomberg News. (2012). "Charities Deceive Donors Unaware Money Goes to a Telemarketer." Website link: <http://www.businessweek.com/news/2012-09-12/charities-deceive-donors-unaware-money-goes-to-a-telemarketer>
- Center for High Impact Philanthropy. "Linking Cost and Impact." Website link: http://www.impact.upenn.edu/about/cost-and-impact/linking_considerations_of_cost_and_impact/
- CGAP (2006) "Aid Effectiveness in Micro finance" Focus Note 35, April
- Charity Navigator. "How Do We Rate Charities." Website link: <http://www.charitynavigator.org/index.cfm?bay=content.view&cpid=1284>
- Ebrahim A. and V. Kasturi Rangan (2010) "The Limits of Non-profit Impact" Harvard Business School Working Paper 10-099, May
- Eccles R., I. Ioannou and G. Serafeim (2011) "The Impact of a Corporate Culture of Sustainability on Corporate Behavior and Performance, *Working Paper* 12-035 Harvard Business School, November 25
- Global Impact Investing Network (2010) "Open letter from the Global Impact Investing Network", December 9
- Global Reporting Initiative (GRI). "Reporting Framework." Website link: <https://www.globalreporting.org/reporting/reporting-framework-overview/Pages/default.aspx>
- Global Reporting Initiative (GRI). "Sustainability Disclosure Database." Website link: <http://database.globalreporting.org>
- Grameen Foundation. Progress of out Poverty. Website link: <http://www.progressoutofpoverty.org>
- Hahn R. and P. Dudley (2007) "How Well Does the US Government Do Cost-Benefit Analysis?" AEI-Brookings Joint Center for Regulatory Studies, Working Paper 04-01, March
- Hirschman A. (1982) "Rival interpretations of Market Society: Civilizing, Destructive or Feeble?" *Journal of Economic Literature*, December

Impact Reporting & Investment Standards (IRIS). (2011) "A Performance Analysis for the Impact Investing Industry." Website link: http://iris.thegiin.org/files/iris/Data_Driven_IRIS_report_final.pdf

Little I. and J. Mirrlees (1990) "Project Appraisal and Planning Twenty Years On" in *Proceedings of the World Bank Annual Conference on Development Economics*, Washington D.C.

J.P.Morgan Global Research (2010) "Impact Investments: an emerging asset class", November 29

Mayer C. (2013) "Firm Commitment" Oxford University Press

Michaelowa K. and Borrmann A. (2004) "The Political Economy of Aid Evaluation: The German Case" Nord-Sued Aktuell Vol. XVIII, No.1

New Profit Inc. "Investment Performance." Website link: <http://www.newprofit.com/cgi-bin/iowa/do/54.html>

Norman W. and C. MacDonald (2003) "Getting to the Bottom of the Triple Bottom Line" in *Business Ethics Quarterly*, March

Ravallion M. (2008) "Evaluation in the Practice of Development" World Bank Policy Research Working Paper 4547, Washington D.C., March

Ravallion M. (2012) "Fighting Poverty one Experiment at a Time: A Review Essay on Abhijit Banerjee and Esther Duflo, Poor Economics" *Journal of Economic Literature* Vol. 50, No.1

Robin Hood Foundation. (2009). "Measuring Success: How Robin Hood Estimates the Impact of Grants." Website link: https://www.robinhood.org/sites/default/files/2009_Metrics_Book.pdf

Schreiner M. and J. Yaron (1999) "The Subsidy Dependence Index and Recent Attempts to Adjust it" submitted to *Savings and Development*, February 2

Tuan M. (2008) "Measuring and/or Estimating Social Value Creation", Bill and Melinda Gates Foundation, December 15

William and Flora Hewlett Foundation. (2008) "Making Every Dollar Count: How expected return can transform philanthropy." Website link: http://www.hewlett.org/uploads/files/Making_Every_Dollar_Count.pdf