

Matthew Hall

Evaluation logics in the third sector

**Article (Accepted version)
(Refereed)**

Original citation:

Hall, Matthew (2014) *Evaluation logics in the third sector*. [Voluntas: International Journal of Voluntary and Nonprofit Organisations](#), 25 (2). pp. 307-336. ISSN 0957-8765

DOI: [10.1007/s11266-012-9339-0](https://doi.org/10.1007/s11266-012-9339-0)

© 2012 [International Society for Third-Sector Research](#) and [The Johns Hopkins University](#)

This version available at: <http://eprints.lse.ac.uk/46365/>

Available in LSE Research Online: August 2014

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

Full text copy of:

Hall, M. Evaluation logics in the third sector. 2014. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 25, 307-336.

Abstract

In this paper I provide a preliminary sketch of the types of logics of evaluation in the third sector. I begin by tracing the ideals that are evident in three well-articulated yet quite different third sector evaluation practices: the logical framework, most significant change stories, and social return on investment. Drawing on this analysis, I then tentatively outline three logics of evaluation: a scientific evaluation logic (systematic observation, observable and measurable evidence, objective and robust experimental procedures), a bureaucratic evaluation logic (complex, step-by-step procedures, analysis of intended objectives) and a learning evaluation logic (openness to change, wide range of perspectives, lay rather than professional expertise). These logics draw attention to differing conceptions of knowledge and expertise and their resource implications, and have important consequences for the professional status of the practitioners, consultants, and policy makers that contribute to and/or are involved in evaluations in third sector organizations.

Introduction

Performance measurement and evaluation is an important and increasingly demanding practice in the third sector (Reed & Morariu, 2010, Benjamin, 2008; Carman, 2007, Eckerd & Moulton, 2011; Charities Evaluation Service, 2008; Bagnoli & Megali, 2010). There is a focus on how to improve the measurement and evaluation of third sector organisations through multi-dimensional frameworks (Bagnoli & Megali, 2010) and balanced scorecards (Kaplan, 2001) and through linkages to strategic decision making (LeRoux & Wright, 2010). Performance measurement and evaluation is also used for a variety of purposes, such as demonstrating accountability to donors and beneficiaries, analyzing areas of good and bad performance, and promoting the work of third sector organizations to potential funders and a wider public audience (Ebrahim, 2005; Barman, 2007; Roche, 1999; Fine et al., 2000; Hoefer et al., 2000; Campos et al., 2011). This wide range of purposes can result in a field that is often cluttered with new ideas, novel approaches and the latest toolkits (Jacobs et al., 2010).

A diversity of approaches can generate debate about the design of particular performance measurement and evaluation techniques and the relative merits of the different types of information that they produce (Charities Evaluation Service, 2008; Reed & Morariu, 2010; Waysman & Savaya, 1997; Eckerd & Moulton, 2011). A common disagreement in such debates concerns the claim that case studies and stories can be ‘subjective’ whereas performance indicators and statistics are more ‘objective’ (e.g., see discussion in Jacobs et al., 2010; Wallace et al., 2007; Abma, 1997; Porter & Kramer, 1999). Of course, in the context of a particular third sector organization with particular stakeholder demands, it is likely that some methods can produce information that is considered more reliable and valid than others. What is of interest here, however, is that such disagreements are likely to reflect (at least in part) a normative belief in the superiority of particular approaches to performance measurement and evaluation, rather than the inherent strengths and weaknesses of any

particular technique. Within the literature, however, there is little explicit analysis of the normative ideals that underpin performance measurement and evaluation practices in third sector organizations (Bouchard, 2009a; 2009b; Eme, 2009). As such, the purpose of this paper is to use an analysis of the ideals evident in different evaluation approaches to develop a tentative sketch of the different logics of evaluation in the third sector. Such an approach can advance understanding of performance measurement and evaluation practice and theory in the third sector in three ways.

First, it directs attention to the normative properties of performance measurement and evaluation approaches by focusing on logics, that is, the broad cultural beliefs and rules that structure cognition and guide decision making in a field (Friedland & Alford, 1991; Lounsbury, 2008; Marquis and Lounsbury, 2007). Under this approach, multiple evaluation logics can create diversity in practice and an argumentative battlefield regarding which evaluation practices are most appropriate (Eme, 2009). Such an approach indicates that seemingly technical debates, such as the relative merits of indicators and narratives, can be manifestations of deeper disagreements about what constitutes the 'ideal' evaluation process. Developing an understanding of evaluation logics is important because the literature on third sector performance measurement and evaluation is under-theorized and lacking conceptual framing (Ebrahim & Rangan, 2011).

Second, it focuses attention on how different evaluation logics can privilege different kinds of knowledge and methods of knowledge generation (Bouchard, 2009a; Eme, 2009). This is critical because evaluations provide an important basis from which third sector organizations seek to establish and maintain their legitimacy in the eyes of different stakeholders (Ebrahim, 2002; 2005; Enjolras, 2009). As such, claims of 'illegitimacy' regarding particular evaluation approaches may be explained, at least in part, by an analysis of conflicting logics of evaluation. For example, stakeholders are increasingly demanding that

evaluation information be quantitative in nature and directed towards demonstrating the impact of third sector organizations (e.g., McCarthy, 2007; LeRoux & Wright, 2010; Benjamin, 2008), an approach that can conflict with techniques focused on dialogue and story-telling. It is here that an analysis of evaluation logics can make explicit the ideals that generate preferences for particular forms of knowledge and information in the evaluation process. This may help stakeholders to negotiate and potentially reconcile differences by balancing or blending different types of information and methods of knowledge generation (e.g., Nicholls, 2009; Waysman & Savaya, 1997), whilst also illuminating the ways in which particular approaches may be fundamentally incompatible.

The third and final implication concerns the role of expertise in the evaluation of third sector organizations. This can have important consequences for the role of the evaluator in the evaluation process, and thus the professional status, knowledge and skills of the practitioners, consultants, and policy makers that are involved in evaluations in third sector organizations. It is also critical to issues of power, because particular conceptions of expertise and ‘valid’ information can serve to elevate the interests of certain actors in third sector organizations whilst disenfranchising others (Ebrahim, 2002; Greene, 1999; Enjolras, 2009). This can be through the creation of evaluation techniques that have a particular exclusionary mystique attached to them (Crewe & Harrison, 1998), for example, the use of words and concepts that are difficult to translate across languages and cultures (Wallace et al., 2007). Differing levels and types of expertise also have resource implications, which is a critical issue for third sector organizations that are increasingly required to use more sophisticated evaluation approaches but with limited (or no) funding for such purposes. In this way, greater understanding of the level and types of expertise advanced by particular evaluation logics is important in illuminating how performance measurement and evaluation can affect whose knowledge and interests are considered more legitimate in third sector organizations.

The remainder of the paper contains three sections. In the next section I outline the analysis of three evaluation techniques, the logical framework, most significant change technique, and social return on investment. Following this, section three draws on this analysis to provide a preliminary sketch of the different types of evaluation logics in the third sector, namely, the scientific, bureaucratic and learning evaluation logics. The fourth and final section discusses the implications of the analysis and concludes the paper.

Analysis of evaluation techniques

I selected three different evaluation techniques using two criteria. The first criterion was that the technique was well articulated and there was evidence of its use (although to varying degrees) within third sector organizations. The second criterion was that the techniques exhibit clear differences in evaluation approaches to create variation in the analysis of evaluation ideals. As such, I use an approach that is akin to purposive sampling, that is, selecting evaluation techniques to maximise variation, rather than seeking to obtain a representative sample from the wider population of evaluation approaches. This approach is consistent with the aim of developing a tentative sketch of different types of evaluation logics in the third sector (rather than an exhaustive catalogue).

Using these two criteria, I chose three techniques for analysis: (1) the Logical Framework approach (LFA), (2) the Most Significant Change (MSC) technique, and (3) Social Return on Investment (SROI). All three techniques are well articulated through an assortment of ‘how to’ guides (detailed below) and there is evidence that the techniques (or those that are very similar) are used within third sector organizations (e.g., see Charities Evaluation Service, 2008; Wallace et al., 2007; Campos et al, 2011; Reed & Morariu, 2010; Carman, 2007; Eckerd & Moulson, 2010; Jacobs et al., 2010; Fine et al., 2000; Hoefer et al., 2000). Whilst the three techniques share some common features, they present quite different

approaches to evaluation, particularly concerning the focus on quantitative and qualitative types of data, the level of technical sophistication, the role for consultants and external experts, and the level of participation by different stakeholders.

The empirical focus is the texts that originally described the techniques, typically in the form of ‘how to’ guides. The focus on ‘how to’ guides is important in examining the techniques in their normative form, for understanding the ways in which the techniques themselves were developed, and for analyzing the core assumptions and epistemologies that underlie the different techniques.

For the LFA, I analysed the text that first explicated the technique, that is, Rosenberg and Posner’s *The Logical Framework: A Manager’s Guide to a Scientific Approach to Design and Evaluation* (hereafter RP, 1979). To analyse the MSC I examined two texts by Dart and Davies, *A Dialogical, Story-Based Evaluation Tool: The Most Significant Change Technique* (hereafter DD, 2003) and *The ‘Most Significant Change’ (MSC) Technique: A Guide to its Use* (hereafter, DD, 2005). Finally, for the SROI, I analysed two texts: the first by the originators of the technique, the Roberts Enterprise Development Fund, entitled *SROI Methodology: Analyzing the Value of Social Purpose Enterprise Within a Social Return on Investment Framework* (hereafter, REDF, 2001) and a second text by the New Economics Foundation (who introduced the SROI technique to the United Kingdom) entitled *Measuring Real Value: a DIY guide to Social Return on Investment* (hereafter, NEF, 2007).¹

I frame the analysis around a set of factors, such as the material outputs of the technique, its origins, the preferred methods of producing knowledge, and the role envisioned

¹ Whilst the analysis is focused upon these texts, I also examined a variety of other guide books and texts for each of the techniques. For the LFA: DFID (2009) *Guidance on using the revised logical framework*; SIDA (2004) *The Logical framework approach*; BOND (2003) *Logical framework analysis*; W.K. Kellogg Foundation (2004) *Logic model development guide*; World Bank (undated) *The Logframe handbook*. For the MSC: Dart & Davies (2003b) *MSC Quick Start Guide*; Clear Horizons (2009) *Quick start guide MSC design*; Davies (1998) *An evolutionary approach to organizational learning*. For the SROI: Olsen & Nicholls (2005) *A framework for approaches to SROI analysis*; NEF (2008) *Measuring value: a guide to social return on investment (SROI) 2nd edition*; NPC (2010) *Social return on investment position paper*; Office of the Third Sector UK (2009) *A guide to social return on investment*.

for outside experts. Table 1 provides a full list of the factors, and the corresponding analysis of each evaluation technique.

<insert Table 1 here>

Logical Framework

The LFA was developed throughout the 1970s as a planning and evaluation tool primarily for use by large bilateral and multi-lateral donor organizations, such as USAID, Department for International Development (UK), the United Nations Development Program and the European Commission. The technique spread to many third sector organizations as they began increasingly to receive funds from these and other donor organizations that used the LFA. At the heart of the LFA is the 4x4 matrix. On the vertical, the project is translated into a series of categories, namely, inputs, outputs, purpose and goal. On the horizontal, each category is described using a narrative summary, objectively verifiable indicators, the means of verification, and the listing of important assumptions.

The principal designers and proponents of the LFA were Rosenberg and Posner of the consulting firm Practical Concepts Incorporated in the United States. The title of their text is revealing as the concepts of logic and science are foregrounded, corresponding to the origins of the LFA, which is derived from “the management of complex space age programs, such as the early satellite launchings and the development of the Polaris submarine” (RP, 1979: 2).

The LFA was developed in response to two perceived problems with existing evaluation approaches. First, they were viewed as unclear and subjective, where “planning was too vague...evaluators could not compare-in an objective manner-what was planned with what actually happened” (RP, 1979: 2). Second, evaluations were sites for disagreement that was considered unproductive: “evaluation was an adversarial process... evaluators ended up using their own judgement as to what they thought were ‘good things’ and ‘bad things’” (RP, 1979: 2).

Given the origins of the LFA, it draws directly on what is labeled the “basic scientific method”, which involves viewing projects as a “set of interlocking hypotheses: if inputs, then outputs; if outputs, then purpose” (RP, 1979: 7). It is these interlocking hypotheses that provide the LFA with some of its most recognizable features. First, the translation of project activities and effects into the categories of inputs, outputs, and purposes (or variations thereof). Second, the explicit outlining of a linear chain of causality amongst these categories. The importance of the chain of causality is highlighted visually in the text with a graphic representation, reproduced below in figure 1:

<insert figure 1 here>

Another focus of the LFA is its concern that evaluation be based on “evidence” (RP, 1979: 3), which takes the form of “Objectively Verifiable Indicators” that are the “means for establishing what conditions will signal successful achievement of the project objectives” (RP, 1979: 19). RP (1979) envisage the role for indicators as follows:

Indicators demonstrate results...we can use indicators to clarify exactly what we mean by our narrative statement of objectives at each of the project levels (RP, 1979: 19).

This quote is revealing as the emphasis on indicators as demonstrating results is viewed as “establishing a “performance specification” such that even skeptics would agree that our intended result has been achieved” (RP, 1979: 21). Here, the role of indicators is effectively to prove to outsiders that results have indeed been achieved. In addition, the emphasis on clarifying objectives relates directly to the desire to avoid “misunderstanding or...different interpretations by those involved in the project” (RP, 1979: 18). This is closely linked to what RP (1979) viewed as one of the key problems with extant evaluation practice; disagreements caused by evaluators using their own judgements. In this sense, indicators are imbued with an objective quality that evaluators do not possess, and hence are viewed as not subject to disagreement.

This approach corresponds to wider discussion in RP (1979) on avoiding conflict, where, in contrast to other evaluation approaches, the LFA:

creates a task-oriented atmosphere in which opportunities, progress and problems that may impede that progress can be discussed constructively. Because the manager knows he is not being held accountable for unrealistic objectives...He does not need to worry that he will be blamed for factors outside his control (RP, 1979: 29-30).

Here, the detailed specification of objectives and indicators has two effects. First, it is seen to remove any possibility for conflict and discussions are thus ‘constructive.’ Second, it provides such clarity that being blamed for issues apparently outside one’s control is no longer possible. This corresponds to a broader view on the evaluation process, where “the evaluation task is simply to collect the data for those key indicators and “evaluate” the project against its own pre-set standards of success” (RP, 1979: 40). The use of quotation marks for the word evaluation is revealing, implying that there is actually very little evaluation required; rather, a simple comparison of actual results to pre-set standards. This reduces the role of the evaluator to somewhat of a fact-checker, and one could easily imagine this role being performed by very junior staff or even a computer. More fundamentally, it suggests that real expertise relates to designing projects, not their evaluation. The expert is the one who conceives of projects, translates activities into the hierarchy of objectives, clarifies cause-and-effect relations, and establishes objectively verifiable indicators. Here, the role for the evaluator is secondary, where they may be ‘called-in’ to advise on the feasibility of data collection and to reduce its cost (RP, 1979: 40).

Most Significant Change

The MSC technique was developed in the 1990s as an approach to evaluating complex social development programs. At the centre of the MSC is the regular collection and interpretation of stories about important changes in the program that are typically prepared by

those most directly involved, such as beneficiaries, clients and field staff. The MSC technique has been used to evaluate programs in both developing and developed economies.²

The principal designers and proponents of the MSC technique were Davies and Dart. Both completed PhD projects concerning the MSC; Davies in the UK with field work focused on a rural development program in Bangladesh, and Dart in Australia with field work focused on the agricultural sector in the state of Victoria.

The MSC technique was developed in response to several perceived deficiencies in existing evaluation practice. First, that evaluations were focused on indicators that are “abstract” (DD, 2003: 140) and do not provide “a rich picture of what is happening [but an] overly simplified picture where organizational, social and economic developments are reduced to a single number” (DD, 2005: 12). The most vivid distinction between approaches is made on the cover page of DD (2005), where a picture shows a man standing opposite a woman and child: the man says “we have this indicator that measures...”, to which the woman replies “let me tell you a story...” (see Figure 2).

<insert figure 2 here>

The contrast is further reinforced through an analogy, which states that “a newspaper does not summarise yesterday’s important events via pages and pages of indicators...but by using news stories about interesting events” (DD, 2005: 16). Finally, the alternative names for the MSC also reveal its ‘anti-indicator’ foundations, which include labels such as “Monitoring-without-indicators” and “The ‘story’ approach” (DD, 2005: 8).

A further perceived problem with existing evaluation approaches is that they focus on examining intended rather than unintended changes. Here, the MSC “does not make use of pre-defined indicators” (DD, 2005: 8), and the criteria for selecting stories of significant

² For further information about the MSC, see <http://mande.co.uk/special-issues/most-significant-change-msc/>, and www.clearhorizon.com.au/flagship-techniques/most-significant-change/. More information, including example reports and applications, can be obtained by joining the MSC Yahoo group: <http://groups.yahoo.com/group/MostSignificantChanges/>.

change “should not be decided in advance but should emerge through discussion of the reported changes” (DD, 2005: 32). This change in approach is most clearly specified in the contrast made between deductive and inductive approaches:

Indicators are often derived from some prior conception, or theory, of what is supposed to happen (deductive). In contrast, MCS uses an inductive approach, through participants making sense of events after they have happened (DD, 2005: 59).

Here the focus of evaluation is quite different to techniques like the LFA, in that there is no comparison to a pre-defined set of objectives. The focus is on identifying, selecting and interpreting stories after events have taken place. The evaluation process itself becomes much more open-ended, particularly as expected outcomes do not frame (at least explicitly) the evaluation process. This is reinforced by guidance that only steps four, five and six (of 10) are essential to the MSC approach, with the possibility of excluding other steps if not deemed necessary by the organizational context and/or reasons for using MSC.

The nature of expertise is also quite different, in that it concerns the development and interpretation of significant change stories after the fact. As such, the evaluator’s task in the MSC is to encourage people to write stories, to help with their selection, and to motivate and inspire others during the evaluation process. This is likely to require skills in narrative writing, facilitation of groups, and interpreting ambiguous events, whereas developing pre-defined indicators is likely to require skills in project design, performance measurement, verification methods, and quantitative data collection.

The shift in the nature of expertise in the MSC technique is also one that “requires no special professional skills” (DD, 2005: 12) and is designed to “encourage non-evaluation experts to participate” (DD, 2003: 140). This is reinforced through the emphasis given to the gathering of stories by field staff and beneficiaries rather than evaluation experts, where the “MSC gives those closest to the events being monitored (e.g., the field staff and beneficiaries) the right to identify a variety of stories that they think are relevant” (DD, 2005: 60).

The MSC is also overtly ‘political’, in the sense that disagreement and conflict are encouraged. For example, the story selection process:

involves considerable dialogue about what criteria should be used to select winning stories...The choice of one story over another reinforces the importance of a particular combination of values. At the very least, the process of discussion involved in story selection helps participants become aware of and understand each other’s values (DD, 2005: 63).

This quote reveals how the surfacing of and discussion about different values is of critical importance in the MSC. In fact, DD (2003: 138) go so far as to argue that the deliberation and dialogue surrounding the selection of stories is the most important part of the MSC technique. Finally, the methods of producing knowledge under the MSC are qualitative in nature, relying on interviews, group discussions, and narrative.

Social Return on Investment

SROI was developed in the 1990s as an approach to analyzing the value created by social enterprises. The principal designer and proponent of SROI was the Roberts Enterprise Development Fund in the USA. They developed the approach to provide an estimate of the social value generated by social enterprises that they had funded. At the centre of the SROI technique is the production of an SROI report, which is envisioned to include a set of SROI metrics, along with organizational data, project descriptions and case studies of participant experiences. SROI has been used predominately in developed economies such as the USA, but has also been introduced to the UK through the New Economics Foundation.

SROI was developed in response to concerns over existing approaches to philanthropy and their associated techniques of evaluation. In particular, REDF (2001: 10-11) distinguish “Transactive Philanthropy” from “Investment Philanthropy”, where the problem with transactive philanthropy is that:

success is defined as the amount of one’s perceived value created in the sector...the number of grants given and by the size of one’s assets. There is often no real connection made between

the dollars one provides third sector organizations and the social value generated from that support (REDF, 2001: 10).

This quote highlights concerns over the definition of success used by ‘transactive’ philanthropists, and contrasts this to an ‘investment philanthropy’ approach that makes much stronger connections between the provision of funds and social value generated. It is here that SROI becomes linked to investment philanthropy and its ideals of long-term value creation. The title of the NEF (2007) text is revealing in this respect, as it refers to “measuring real value”, giving the impression that it is only through the use of SROI that the true effects of projects are captured.

The origins of the SROI are rooted in private sector evaluation approaches and their associated techniques, where “special attention is given to the application of traditional, for-profit financial metrics to non-traditional, nonprofit, social purpose enterprises” (REDF, 2001: 7). These traditional financial metrics include “standard investment analysis tools...discounted cash flow...net present value analysis” (REDF, 2001: 14). The private sector linkages also extend to the presentation of the analysis, where “SROI reports are similar to for profit company stock reports” (REDF, 2001: 13).

Given the application of financial metrics, a key focus of SROI is quantifying, in financial terms, the outcomes of the organization and/or its projects. In particular, the original formulation of the SROI required the calculation of six SROI metrics:

These first three metrics measure what a social purpose enterprise is “returning” to the community. The next three metrics compare these returns against the philanthropic investments required to generate them. This comparison of returns generated to investments is articulated in the Index of Return...An Index greater than one shows that excess value is generated. If the Index is less than one, value is lost (REDF, 2001: 18).

Here, the SROI metrics make two connected translations. First, the effects of the activities of a social enterprise are expressed in a financial return. Second, this measure of return, along with a measure of investment, is used to make a further translation into an index. It is at this

stage that the index is used to make an evaluation of the enterprise and/or project, that is, whether it lost or created value, as illustrated in Figure 3 below:

<insert figure 3 here>

Whilst both the REDF (2001) and NEF (2007) guides show the mechanics of these translations in some detail, actually conducting such analysis does require certain skills, such as numeracy, financial literacy, an affinity with Excel, and an understanding of concepts such as the time value of money and discounted cash flows. Skills in planning are also required, with the SROI approach following a sequence of steps. For example, NEF (2007) present the 10 stages of an SROI analysis, with the end of most stages requiring the completion of ‘checklists’ to ensure that the stage is properly completed before proceeding to the next stage in the process.

In the SROI approach there is also a concern with providing contextual information to support interpretation of numerical data, highlighted in the guidance that “your final report should comprise much more than the social returns calculated...[you should use] supplemental information such as participant surveys and other data that help to convey the story behind the results” (NEF, 2007: 53). This quote is revealing, however, in that such data is viewed as ‘supplemental’ and being ‘behind the results’, and thus gives the clear impression that it is secondary to the primary focus on metrics.

The SROI also seeks to adjust outcomes to reflect the influence of outside factors. For example, outcomes of projects should account for “attribution” (NEF, 2007: 27) and be adjusted further for “deadweight” (NEF, 2007: 27), that is, the “extent to which the outcomes would have happened anyway” (NEF, 2007: 27). This resonates with the use of a ‘control group’ to ensure that the observed effects were indeed caused by the project (i.e., the ‘treatment condition’) and not exogenous factors.

The SROI also employs specific terminology to describe the approach itself. In particular, NEF (2007: 10) state, “we use the word ‘study’ to describe the process of preparing an SROI report and we refer to the person doing the work as the ‘SROI researcher’”. Such a researcher is also imbued with particular characteristics, such as “independence and objectivity” (NEF, 2007: 36). In general, such terminology shows a concern with presenting the SROI in a particular light, specifically that of an independent and objective research study.

Outline of evaluation logics

Drawing on the above analysis, in this section I provide a preliminary sketch of the types of logics of evaluation in the third sector, i.e., the broad cultural beliefs and rules that structure cognition and shape evaluation practice in third sector organizations (c.f., Friedland & Alford, 1991; Lounsbury, 2008; Marquis and Lounsbury, 2007). The evaluation logics were developed in three steps. In the first step I used the above analysis of the LFA, MSC and SROI approaches to identify evaluation ideals. Steps two and three sought to broaden the analysis beyond these specific evaluation approaches. In the second step I analysed a broader range of evaluation practices, such as the scorecards (e.g., Kaplan, 2001), outcome frameworks (e.g., Urban Institute, 2006), participatory methods (e.g., Keystone Accountability, undated) and expected return methods (Acumen Fund, 2007). In the third and final step I examined writings on evaluation practice in the third sector, drawn from the third sector, evaluation and social development literatures, to locate discussion of and or reference to these evaluation ideals.³ Overall, this analysis resulted in the development of three logics of evaluation: a scientific evaluation logic, a bureaucratic evaluation logic and a learning

³ The academic journals examined included *Nonprofit and Voluntary Sector Quarterly*, *Voluntas*, *Nonprofit Management and Leadership*, *American Journal of Evaluation*, *Evaluation*, *Public Administration and Development*, and the *Journal of International Development*.

evaluation logic.⁴ In the analysis that follows, I develop a set of ideals that are characteristic of each evaluation logic, provide linkages to prior literature, as well as examples from various evaluation techniques that illustrate the ideals. To help compare and contrast each evaluation logic, the ideals are grouped according to one of four questions: what makes a ‘quality’ evaluation, what characterizes a ‘good’ evaluation process, what is the focus of evaluation, and what is the role of the evaluator. A summary of each evaluation logic, its ideals and associated examples is provided in Table 2.

<insert Table 2 here>

Scientific evaluation logic

The scientific evaluation logic echoes the scientific method, with a strong focus on systematic observation, gathering of observable and measurable evidence, and a concern with objective and robust experimental procedures. Its ideals are those of proof, objectivity, anti-conflict and reduction, and the evaluator’s role is that of a scientist.

Of fundamental concern to the scientific evaluation logic is that the evaluation process is focused on establishing ‘proof’, that is, that the claims made about the effects of projects must be demonstrated by the use of evidence, and that alternative explanations for those effects have been considered and ruled out. Surveys show that third sector organizations are increasingly being asked to provide proof of causality and attribute outcomes to specific interventions (Charities Evaluation Service, 2008), and to gather data that shows the concrete, tangible changes that have resulted from the support of foundations (Easterling, 2000). These concerns are evident in the LFA, where verifiable indicators provide evidence of project effects. In the SROI, there is explicit concern with taking account of external influences via the concept of ‘attribution.’ Furthermore, echoing the use of a control group in randomized

⁴ Here the term ‘scientific’ functions as a descriptive label to characterise a particular approach to evaluation. The use of this term makes no claims about the scientific merits or value of this approach relative to other evaluation approaches.

trials, the concept of ‘deadweight’ demands consideration of what effects would have transpired in the event that the project was not conducted. Similarly, Acumen’s Best Available Charitable Option (BACO) ratio requires discounting social impact to that which can be ‘credited specifically to Acumen’s financing’ (Acumen Fund, 2007: 3).

Along with proof, evaluative judgements and data collection processes must be ‘objective’ under a scientific evaluation logic. That is, they are not influenced by the personal feelings or preferences of the evaluator or other participants in the evaluation process. This resonates with Fowler (2002), who argues that under a ‘hard’ science approach, attempts to evaluate NGO performance are characterized as objective, in the sense that knowledge generated is independent of the persons doing the observing. Similarly, Blalock (1999, 139) states that “scientific evaluations” are those that are carried out by researchers independent of the program being studied. Here, evaluators embody a professional expertise characterized by detachment and scientific rigor (Marsden & Oakley, 1991). In the context of specific evaluation approaches, the LFA stresses that indicators must be verified objectively, and SROI states that evaluators should be independent and objective researchers.

Under a scientific evaluation logic an ideal evaluation process is also ‘anti-conflict.’ Here, evaluations should be designed and conducted so as to avoid, as far as possible, any conflict amongst evaluators and others involved in the process. There is a belief that evaluation methods can be value-neutral and that they should de-emphasize or ignore the political processes involved in evaluation (Marsden & Oakley, 1991; Jacobs et al., 2010). This links strongly to ideals of objectivity, in that one way to avoid conflict is to ensure the objectivity of data and of evaluators. The LFA exemplifies the ideal of ‘anti-conflict’, with its explicit desire to move away from evaluation as an adversarial process.

A further ideal of the scientific evaluation logic is to be ‘reductive.’ That is, the tendency of an evaluation approach to represent project outcomes in a simplified form. Being

reductive involves the use of maps, models, matrices, and numerical calculations to represent objects in social development projects (Crewe & Harrison, 1998; Scott, 1998). A fundamental feature of the LFA is the construction of a 4x4 matrix to represent the project and its outcomes, which provides a short and convenient summary of a project, simplifying complex social situations and making them relatively easy to understand (Jacobs et al., 2010). In SROI, whilst there is concern with presenting case studies and background data, the core of the approach is the calculation of metrics and indices to represent the social return of the project. The use of monetization in SROI can reduce the complex information about third sector organizations into data that can easily be compared and valued (Lingane & Olsen, 2004). The focus on reduction is also evident in several other techniques, such as the production of a scorecard using the balanced scorecard (Kaplan, 2001), the calculation of the BACO ratio (Acumen Fund, 2007) and the development of an outcome sequence chart in the Common Outcomes Framework (Urban Institute, 2006).

Finally, under a scientific evaluation logic the evaluator is conceived of as a ‘scientist,’ that is, an actor who conducts experiments, engages in research and reports findings. This is exemplified in the LFA, where evaluators use a ‘scientific approach’, and in SROI, where evaluators take on the role and label of ‘researcher.’

Bureaucratic evaluation logic

The bureaucratic evaluation logic is rooted in ideals of rational planning, with a strong focus on complex, step-by-step procedures, the limiting of deviations from such procedures, and analysis of the achievement of intended objectives. Its ideals are those of categorization, sequential, intended effects, and hierarchy, and the evaluator’s role is that of an implementer.

An ideal of the bureaucratic evaluation logic is that of ‘categorization.’ The most important feature of categorization is not so much the use of categories per se, but that those

categories are pre-defined and thus standardized across all evaluations. In effect, categorization means that the project is mapped to the demands of the evaluation approach, rather than each project creating a (at least somewhat) unique set of categories. This resonates with the way in which categorization is considered a key social process of bureaucracy (Stark, 2009). The LFA exemplifies the ideal of categorization, with project activities translated into four pre-defined categories: inputs, outputs, purpose and goal. The SROI also resonates with the categorization ideal, as the value generated by projects is translated into three pre-existing forms: economic value, socio-economic value and social value. Categorization is also a feature of several other techniques, such as the Balanced Scorecard with its use of four pre-defined perspectives, the development of a set of common outcome indicators in the Common Outcome Framework (Urban Institute, 2006), and the translation of social value into the categories of total output and social impact in determining the BACO ratio (Acumen Fund, 2007).

With its roots in rational planning, the bureaucratic evaluation logic emphasizes a ‘sequential’ evaluation process. That is, evaluations should proceed according to a step-by-step process, and, more importantly, it is necessary to complete each stage before proceeding to the next, which corresponds to a perspective on social development as a linear process (Crewe & Harrison, 1998; Wallace et al., 2007; Howes, 1992). In the context of SROI, NEF (2007) presents the 10 stages of an SROI analysis, with each stage following from the completion of all prior stages, and evaluators are provided with ‘checklists’ to ensure that stages are complete before moving to the next stage. In contrast, the MSC approach states that only steps 4, 5 and 6 (of 10) are essential, with the possibility of excluding other steps if not deemed necessary by the organizational context and/or reasons for using MSC.

A bureaucratic evaluation logic also privileges the analysis of ‘intended effects,’ i.e., whether or not the effects of the project that were envisioned prior to its completion did in

fact eventuate. Here, evaluation is focused on the attainment of pre-determined goals (Howes, 1992) and is often associated with a downgrading of the achievement of unintended effects, whether good or bad (Gasper, 2000). The focus on a limited set of outcomes can mean that the true complexity of a program is frequently ignored in the information production process (Blalock, 1999). The LFA exemplifies this ideal, with its establishment of objectives and indicators at the planning stage of projects in order to compare actual outcomes with those plans.

‘Hierarchy’ is a further ideal of the bureaucratic evaluation logic. This ideal involves the creation of a ranking amongst aspects of the evaluation process, such as project activities, results of the project and/or the types of information to be used in the evaluation, such that some features are considered higher than and/or more important than others. In the context of specific evaluation approaches, ‘hierarchy’ can be explicit or implicit. For example, in the LFA, there is an explicit hierarchy of project objectives, where achieving what is intended at one level leads to the next one higher up and so on until the final and ultimate goal is reached (Gasper, 2000). In the SROI, there is an implicit hierarchy of forms of data, whereby participant surveys and other qualitative data are ‘supplemental’ to other more quantitative forms of data. The Balanced Scorecard also creates a hierarchy, with the ‘overall mission’ at the ‘top’ of the scorecard (Kaplan, 2001), and the Impact Planning, Assessment and Learning method creates a ‘pyramid of indicators’ with ‘high-level outcome indicators’ at the top and ‘local programme level indicators’ at the bottom (Keystone, undated).

Finally, the ideal evaluator under a bureaucratic logic is that of the ‘implementer.’ The evaluator’s role is to ensure that the evaluation proceeds, as far as possible, according to the specified methodology. The evaluator is limited to the collection of data, providing a technically competent and politically neutral expert in order that appropriate information is provided to the decision makers (Abma, 1997). The evaluator in an LFA is responsible for

ensuring that activities are correctly categorized and that indicators are both objective and verifiable. In the SROI, the evaluator ensures that the stages of analysis are adhered to and that all necessary activities have been undertaken via the use of checklists.

Learning evaluation logic

The learning evaluation logic privileges an openness to change and the unexpected, the incorporation and consideration of a wide ranges of views and perspectives, and a focus on lay rather than professional expertise. Its ideals are those of richness, belief revision, and egalitarianism, and the evaluator's role is that of a facilitator.

An important ideal of the learning evaluation logic is that of 'richness,' where the evaluation process privileges analysis of the fullest possible range of and variation in project effects. Here, evaluation and the social development process recognize the importance of narratives, which are never final products but are always in a state of 'becoming' (Conlin & Stirrat, 2008). Similarly, richness invites a rejection of performance measures that can capture but only a small fraction of what is important and invites a focus on the overall story, textures and nuances that can reveal the multiple levels of human experience (Greene, 1999). Such an approach helps to guard against so-called 'context-stripping,' that is, evaluation as though context did not exist but only under carefully controlled conditions (Guba & Lincoln, 1989). The central feature of the MSC approach exemplifies the ideal of 'richness,' with its focus on telling stories of events and providing what are called 'thick descriptions.' Richness is also clearly evident in the Impact Planning, Assessment and Learning method, where it recognises that change can be 'uncertain and unpredictable' and that the impact of third sector organizations can be 'intended or unintended, positive or negative – and often both together' (Keystone, undated: 3). To some extent, the SROI approach also attempts to convey a rich picture through the production of an SROI report and case studies of participants'

experiences, and the Balanced Scorecard focuses on providing a ‘multidimensional framework for measuring and managing nonprofit effectiveness’ (Kaplan, 2001: 357).

‘Belief revision’ is a further ideal of the learning evaluation logic, where evaluation is focused on uncovering the unexpected and on searching for deviations. The object of belief revision can take several forms, such as how the project was conducted and looking for ways to improve, examining what didn’t work, what was not achieved, and why, and analysis of the evaluation process itself and how it could be refined. In this way, the ideal of belief revision has an outlook that is orientated towards the future, where analysis of what has already happened is premised on making changes to future plans and activities. Here, the evaluation process can construct self-reflective moments that allow individuals to examine the realities they confront, which creates an atmosphere of continual learning (Campos et al., 2011).

Similarly, evaluation can involve generating knowledge to learn and change behavior, whereby evaluation systems help to improve programs by examining and sharing successes as well as failures through engagement with stakeholders at all levels (Ebrahim, 2005). Aspects of the MSC approach are focused squarely on belief revision, such as the emphasis on identifying unexpected changes that result from projects, and an explicit concern with analyzing how the MSC process itself could be revised. The Impact Planning, Assessment and Learning method also embodies belief revision, with its focus on developing ‘learning relationships’ and the use of reflection to examine ‘what is working, why it is working, what must be sustained and what must be changed’ (Keystone, undated: 3 & 10). Similarly, the Common Outcome Framework suggests that ‘outcome data should be used to identify where results are going well and where not so well...this process is what leads to continuous program learning’ (Urban Institute, 2006: 15). The SROI approach also embodies elements of this ideal, with a concern on using SROI analysis as a means to learn more about how projects contribute to social value creation.

The learning evaluation logic is also ‘egalitarian’ in nature, with an explicit concern in ensuring that the evaluation process and its associated techniques are readily understandable to a wide range of stakeholders. There should be minimal (if any) need for training in specialized techniques or abstract concepts, and experts are not required. The MSC approach is, to a large degree, built on the egalitarian ideal, with its use of story-telling, a widely used and even intrinsically human practice, and the insistence that the approach itself should require no special professional skills. The Impact Planning, Assessment and Learning method also focuses on the use of dialogue techniques and ‘change journals in which staff record the informal feedback and changes that they observe in their daily work’ (Keystone, undated: 9). In contrast, the language and approach of techniques like the LFA are often experienced by field staff as alienating, confusing and culturally inappropriate (Wallace et al., 2007). In addition, the use of techniques like random assignment and matched comparison groups are exceedingly difficult to implement within most third sector settings, particularly given the level of resources usually available for evaluation (Easterling, 2000). This is evident in the Common Outcome Framework whereby there is criticism of ‘classic program evaluation’ and its focus on ‘lengthy academic evaluations and complex, meaningless statistical analyses’ (Urban Institute, 2006: 3).

The ideal evaluator under a learning evaluation logic is that of the ‘facilitator.’ The evaluator’s role is to help others to participate in the evaluation process, and to make their tasks as easy as possible. This role has strong links to wider participation and empowerment discourses, whereby outsiders with ‘expert’ technical skills should relinquish control and serve rather to facilitate a process of learning and development (Howes, 1992). Under this perspective, the evaluator works actively to develop a transactional relationship with the respondents on the basis of participation (Abma, 1997) and can prepare an agenda for negotiation between stakeholders, taking on the role of a moderator (Guba & Lincoln, 1989).

In the MSC approach, the evaluator's task is to encourage people to write stories, to help with their selection, and to motivate and inspire others during the evaluation process. In the Impact Planning, Assessment and Learning method there is also a focus on facilitating the involvement of constituents, and flexibility is emphasized with organizations encouraged to use the method in a way that 'suits their needs and context' (Keystone, undated: 4).

Discussion

In this paper I have provided a preliminary sketch of the types of logics of evaluation in the third sector, namely, the scientific, bureaucratic and learning evaluation logics. The evaluation logics are akin to 'ideal-types' (Weber, 1904/1949) in that they serve to highlight the types of beliefs and rules that structure the practice of evaluation in the third sector. In this way, although the evaluation logics were developed from analysis of particular evaluation practices, they will not correspond to all the features of any particular evaluation approach. Indeed, a specific evaluation practice can align with ideals from different evaluation logics, as is shown in Table 2, where characteristics of SROI align with ideals from each of the three evaluation logics.

Analysis of different evaluation logics indicates that many debates, such as conflicts over the use of different forms of data (quantitative vs. qualitative data being a common one) are, at least to some extent, manifestations of disagreements about what constitutes an ideal evaluation process. This is important because many ideals of the three evaluation logics sketched here may be potentially incompatible, and it is these situations that are rife for contestation over which evaluation practices are most appropriate.

The scope for conflict becomes particularly evident through a comparison of how the ideals of the different evaluation logics address the four evaluation questions (see Table 2). In relation to what makes a quality evaluation, the logics differ considerably in their responses.

A scientific evaluation logic values ‘proof’ and accounting for alternative explanations, a bureaucratic logic values mapping projects to pre-defined categories, and a learning evaluation logic values ‘richness’ and analysis of variation in project effects. Here, the ideals ‘proof’ and ‘richness’ may be compatible, whereby an evaluation approach employs both indicators and case studies, such as SROI’s attempt to combine a focus on metrics with case studies of participants’ experiences. However, SROI has been criticised for overly focusing on financial value at the expense of a fuller and more rounded understanding of project effects (Durie, Hutton & Robbie, 2007). Categorization appears fundamentally incompatible with richness, as it is the translation of activities and effects into pre-defined categories that tends to preclude analysis of their nuances and variations, a criticism that has been made of the LFA (Howes, 1992).

The ideals that characterize a ‘good’ evaluation process also differ, where a scientific evaluation logic values an ‘objective’ process that is free of conflict, a bureaucratic evaluation logic values a ‘sequential’ process, whereas a learning evaluation logic is concerned that the process be ‘egalitarian.’ The use of dialogue and story-telling in particular is likely to generate conflict, because although its requirement of little expertise resonates with the egalitarian ideal, such a practice is likely to clash with the ideal of objectivity because stories are subjective experiences. This is evident in criticisms directed at the MSC regarding the ‘bias’ towards good-news stories (Dart & Davies, 2003). There have been attempts to combine a focus on sequential processes with an egalitarian framework, such as linking elements of the logical framework with greater stakeholder involvement in project planning and selection of project objectives (e.g., SIDA, 2006). However, this seems difficult in practice, for example, where the LFA is typically criticised for placing the evaluation expert in a privileged position (Howes, 1992).

Regarding the focus of evaluation, a scientific evaluation logic is aimed at reducing and simplifying outcomes, a bureaucratic evaluation logic values both an analysis of intended effects coupled with the creation of hierarchies, and a learning evaluation logic is concerned with belief revision. In some respects, these ideals are compatible in that being reductive need not preclude belief revision, as illustrated in the SROI approach whereby SROI ratios can be used to learn more about social value creation. Conversely, techniques like MSC have been criticised for not being able to produce summary information to judge the overall performance of a programme (Dart & Davies, 2003). Reconciling the ideals of intended effects and belief revision may also be difficult, for example, as the LFA has been criticised for creating a 'lock-frame' that blocks opportunities for learning and adaptation (Gasper, 2000).

An understanding of these different evaluation logics thus reveals that they privilege different kinds of knowledge and the desired process for knowledge generation. This highlights the role of different epistemologies in debates about the merits of different evaluation practices, and thus recognition of the way in which different evaluation logics can have important implications for whose knowledge and interests are considered more legitimate in third sector organizations (c.f., Ebrahim, 2002; Greene, 1999). This is critical in an arena where evaluations provide an important basis from which third sector organizations seek to establish and maintain their legitimacy in the eyes of different stakeholders (Ebrahim, 2002; 2005; Enjolras, 2009).

Implications and conclusion

This study has shown that developing an understanding of evaluation logics is important given the lack of theorization and conceptual framing in research on performance measurement and evaluation in the third sector (Ebrahim & Rangan, 2011). An examination of evaluation logics helps to go beyond a first-degree level understanding of evaluation

techniques by highlighting the normative properties of different evaluation approaches. To this end, the paper contributes to the literature by providing a framework that can be used to dissect both existing and proposed evaluation techniques. As shown in the paper, Table 1 can be used to understand the characteristics of different evaluation techniques according to a set of common factors, such as the material outputs, the stated purpose of the evaluation technique, and the type and extent of expertise. The three evaluation logics and their associated ideals, as outlined in Table 2, can be used to analyze the normative properties of specific evaluation techniques. Collectively, this framework can help to develop a more conceptual analysis of evaluation practices in the third sector.

The framework also has implications for disputes that can arise over different evaluation approaches. In particular, the evaluation logics can help to determine whether debates over the merits of particular evaluation approaches (e.g., use of qualitative and quantitative data or subjective/objective procedures) are more methodological in nature (e.g., disagreements about the validity with which particular data was collected and analysed) or driven more by a particular evaluation ideal or position (e.g., a belief in the superiority of particular forms of data or methods of data collection). The ability to differentiate better between methodological and ideological critiques may go some way towards exposing the nature of the viewpoints advanced by particular evaluation techniques and/or experts, and thus whether such disagreements can be resolved.⁵ For example, a methodological critique could be addressed by changing the evaluation technique to improve its validity, whereas an ideological critique is more likely to prove intractable even in the face of adaptations to the evaluation methodology.

The analysis of different evaluation logics also reveals that they can have important practical implications. In particular, generating evaluations that can accommodate the ideals

⁵ I thank an anonymous reviewer for suggesting this line of reasoning.

of different stakeholders may be very difficult, and require careful design such that, even within a single evaluation, the ideals of different evaluation logics can be accommodated as far as possible. Alternatively, recognition of different logics of evaluation can also provide scope for third sector organizations to ‘push back’ against demands for specific types of evaluation, particularly those from funders, by highlighting that the evaluation logics of such demands are potentially fundamentally inconsistent with evaluation logics embodied in established evaluation practices.

A related practical consideration is that of cost, particularly pertinent when funding for evaluation activities is limited or non-existent. Although often not explicitly addressed, the resource implications that stem from different evaluation logics are likely to be of critical importance to third sector managers, and, indirectly, for funders themselves. For example, conducting evaluations to satisfy the ideal of ‘proof’ under the scientific evaluation logic has been recognised as complex and expensive, as reflected in the World Bank’s development of guidance to conduct impact evaluations in the presence of budget, time and data constraints (World Bank, 2006). In contrast, more ‘egalitarian’ approaches are likely to require less expertise and thus, from a cost perspective, can have distinct advantages. These differing cost implications of the evaluation logics can, at least in part, be traced to the origins of specific evaluation techniques. For example, elements of the scientific and bureaucratic evaluation logics are located in techniques typically developed by funders and donors (e.g., USAID and the LFA, REDF and SROI), a setting in which the required money and expertise is perhaps more readily available and forthcoming. In contrast, techniques that map very closely to the learning evaluation logic, like the MSC, were developed in and for third sector organizations themselves (e.g., see Davies (1998)), and, perhaps not surprisingly, appear more carefully attuned to the demands that evaluation techniques can place on resources and expertise. In general, an analysis of evaluation logics can help to reveal the differing resource implications

that can follow from advocating (or even requiring) particular approaches to evaluation in third sector organizations.

A final practical implication concerns the knowledge and skills required of evaluators. An evaluator as scientist is likely to require knowledge of the scientific method and skills in quantitative data collection, experimental procedures and the writing up of research findings, whereas an evaluator as facilitator needs knowledge of forms of qualitative inquiry and skills in communication, interpersonal interactions and mediation between groups with different interests/values. This can have important consequences for the professional status of the practitioners, consultants, and policy makers that contribute to and/or are involved in evaluations in third sector organizations.

Using an analysis of the LFA, SROI and MSC evaluation techniques, this study developed three different logics of evaluation in the third sector. This raises several questions for future research. The first question concerns the way in which these evaluation logics relate to evaluation practice, that is, to what extent are the three evaluation logics illustrative of other evaluation techniques in the third sector? A second and related question is: are there other logics of evaluation in the third sector? A third and more broad-ranging question is: how do evaluation logics in the third sector relate to evaluation logics more generally? In this regard, the three evaluation logics appear to relate to more general evaluation approaches. For example, the scientific and bureaucratic evaluation logics have affinity with what Guba and Lincoln (1989) characterise as first (measurement) and second (description) generation evaluation approaches, whereas the learning evaluation logic appears to resonate with what they term 'fourth generation evaluation.' Future research could fruitfully explore these connections.

Through an analysis of specific evaluation techniques, this study has sought to highlight the normative ideals of different evaluation approaches, and to provide a

preliminary sketch of the types of evaluation logics in the third sector. There is much scope for further research to refine the conceptualization of the logics and/or develop additional ones, to examine the ways in which other evaluation techniques reflect and seek to reconcile potentially conflicting evaluation ideals, and to consider the important role that different evaluation logics can play in promoting or discrediting particular types of evaluation information and expertise in third sector organizations.

References

- Abma, T. A. (1997). Playing with/in plurality: revitalizing realities and relationships in Rotterdam. *Evaluation*, 3, 25-48.
- Acumen Fund (2007). *The Best Available Charitable Option*.
http://www.acumenfund.org/uploads/assets/documents/BACO%20Concept%20Paper%20final_B1cNOVEM.pdf, Accessed 19 September 2011.
- Bagnoli, L., & Megali, C. (2011). Measuring performance in social enterprises. *Nonprofit and Voluntary Sector Quarterly*, 40: 149-165.
- Barman, E. (2007). What is the bottom line for third sector organizations? A history of measurement in the British voluntary sector. *Voluntas*, 18, 101-115.
- Benjamin, L. M. (2008). Account space: how accountability requirements shape nonprofit practice. *Nonprofit and Voluntary Sector Quarterly*, 37: 201-223.
- Blalock, A. N. (1999). Evaluation research and the performance management movement: from estrangement to useful integration? *Evaluation*, 5, 117-149.
- BOND. (2003). *Logical Framework Analysis*. Guidance Notes No.4.
- Bouchard, J. M. (2009a). The worth of the social economy. In Bouchard, J. M. *The Worth of the Social Economy: An International Perspective*. P.I.E Peter Lang, Brussels, 11-18.
- Bouchard, J. M. (2009b). The evaluation of the social economy in Quebec, with regards to stakeholders, mission and organizational identity. In Bouchard, J. M. *The Worth of the Social Economy: An International Perspective*. P.I.E Peter Lang, Brussels, 111-132.
- Campos, J. G., Andion, C., Serva, M., Rossetto, A., & Assumepe, J. (2011). Performance evaluation in non-governmental organizations (NGOs): an analysis of evaluation models and their applications in Brazil. *Voluntas*, 22, 238-258.
- Carman, J. G. (2007). Evaluation practice among community-based organizations: research into the reality. *American Journal of Evaluation*, 28: 60-75.
- Charities Evaluation Service. (2008). *Accountability and Learning: Developing Monitoring and Evaluation in the Third Sector*.
- Clear Horizons. (2009). *Quick Start Guide: MSC Design*.
- Conlin, S., & Stirrat, R. L. (2008). Current challenges in development evaluation. *Evaluation*, 14, 193-208.
- Crewe, E. & Harrison, E. (1998). *Whose Development? An Ethnography of Aid*. Zed Books, London.
- Dart, J., & Davies, R. (2003a). A Dialogical, Story-Based Evaluation Tool: The Most Significant Change Technique. *American Journal of Evaluation* 24, 137-155.
- Dart, J., & Davies, R. (2003b). *Quick-Start Guide: A Self-help Guide for Implementing the Most Significant Change Technique (MSC)*.
- Davies, R. (1998). An evolutionary approach to organisational learning: an experiment by an NGO in Bangladesh in Mosse, D., Farrington, J. & Few, A. *Development as Process: Concepts and Methods for Working with Complexity*, Routledge: London, 68-83.
- Davies, R. and Dart, J. (2005). *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*.
- Department for International Development (DFID). (2009). *Guidance on Using the Revised Logical Framework*. DFID Practice Paper.
- Durie, S., Hutton, E. & Robbie, K. (2007). *Investing in Impact: Developing social return on investment*. <http://www.socialimpactsotland.org.uk/media/1200/SROI-%20investing%20in%20impact.pdf>, Accessed 14 September 2012.
- Easterling, D. (2000). Using outcome evaluation to guide grantmaking: theory, reality, and possibilities. *Nonprofit and Voluntary Sector Quarterly*, 29, 482-486.
- Ebrahim, A. (2002). Information struggles: the role of information in the reproduction of

- NGO-funder relationships. *Nonprofit and Voluntary Sector Quarterly*, 31, 84-114.
- Ebrahim, A. (2005). Accountability myopia: losing sight of organizational learning. *Nonprofit and Voluntary Sector Quarterly*, 34, 56-87.
- Ebrahim, A., & Rangan, V. K. (2011). Performance measurement in the social sector: a contingency framework. Social Enterprise Initiative, Harvard Business School, working paper.
- Eckerd, A., & Moulton, S. (2011). Heterogeneous roles and heterogeneous practices: understanding the adoption and uses of nonprofit performance evaluations. *American Journal of Evaluation*, 32: 98-117.
- Eme, B. (2009). Miseries and worth of the evaluation of the social and solidarity-based economy: for a paradigm of communicational evaluation. In Bouchard, J. M. *The Worth of the Social Economy: An International Perspective*. P.I.E Peter Lang, Brussels, 63-86.
- Enjolras, B. (2009). The public policy paradox. Normative foundations of social economy and public policies: which consequences for evaluation strategies? In Bouchard, J. M. *The Worth of the Social Economy: An International Perspective*. P.I.E Peter Lang, Brussels, 43-62.
- Fine, A. H., Thayer, C. E., & Coghlan, A. T. (2000). Program evaluation practice in the nonprofit sector. *Nonprofit Management and Leadership*, 10(3), 331-339.
- Fowler, A. (2002). Assessing NGO performance: Difficulties, dilemmas and a way ahead in Edwards, M. & Fowler, A. *The Earthscan Reader on NGO Management*, Earthscan: London, 293-307.
- Friedland, R., & Alford, R. R. (1991). Bringing society back in: symbols, practices, and institutional contradictions in Powell, W. W. & DiMaggio, P. J, *The New Institutionalism in Organizational Analysis*, Chicago: University of Chicago Press, 232-266.
- Gaspar, D. (2000). Evaluating the 'logical framework approach': towards learning-oriented development evaluation. *Public Administration and Development*, 20, 17-28.
- Greene, J. C. (1999). The inequality of performance measurements. *Evaluation*, 5, 160-172.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Sage Publications, Newbury Park.
- Hoefer, R. (2000). Accountability in action? Program evaluation in nonprofit human service agencies. *Nonprofit Management and Leadership*, 11(2), 167-177.
- Howes, M. (1992). Linking paradigms and practice: key issues in the appraisal, monitoring and evaluation of British NGO projects. *Journal of International Development*, 4, 375-396.
- Jacobs, A., Barnett, C., & Ponsford, R. (2010). Three approaches to monitoring: feedback systems, participatory monitoring and evaluation and logical frameworks. *IDS Bulletin*, 41, 36-44.
- Kaplan, R. S. (2001). Strategic performance measurement and management in third sector organizations. *Nonprofit Management & Leadership*, 11(3), 353-371.
- Keystone (undated). *Impact Planning, Assessment and Learning*.
http://www.keystoneaccountability.org/sites/default/files/1%20IPAL%20overview%20and%20service%20offering_0.pdf, Accessed 19 September 2011.
- LeRoux, K., & Wright, N. S. (2011). Does performance measurement improve strategic decision making? Findings from a national survey of nonprofit social service agencies. *Nonprofit and Voluntary Sector Quarterly*, in press.
- Lingane, A., & Olsen, S. (2004). Guidelines for social return on investment. *California Management Review*, 46, 116-135.
- Lounsbury, M. (2008). Institutional rationality and practice variation: new directions in the institutional analysis of practice. *Accounting, Organizations and Society*, 33, 349-361.

- Marquis, C., & Lounsbury, M. (2007). Vive la resistance: competing logics and the consolidation of U.S. community banking. *Academy of Management Journal*, 50, 799-820.
- Marsden, D., & Oakley, P. (1991). Future issues and perspectives in the evaluation of social development. *Community Development Journal*, 26, 315-328.
- McCarthy, J. (2007). The ingredients of financial transparency. *Nonprofit and Voluntary Sector Quarterly*, 36, 156-164.
- New Economics Foundation (NEF) (2007). *Measuring Real Value: a DIY Guide to Social Return on Investment*.
- New Economics Foundation (NEF) (2008). *Measuring Value: a Guide to Social Return on Investment (SROI)*. 2nd Edition.
- New Philanthropy Capital (NPC). 2010. *Social Return on Investment*. Position paper.
- Nicholls, A. (2009). We do good things, don't we? 'Blend value accounting' in social entrepreneurship. *Accounting, Organizations and Society* 34, 755-769
- Office of the Third Sector (UK Cabinet Office). (2009). *A Guide to Social Return on Investment*.
- Olsen, S., & Nicholls, J. (2005). *A Framework for Approaches to SROI Analysis*.
- Porter, M. E., & Kramer, M. R. (1999). Philanthropy's new agenda: creating value. *Harvard Business Review*, November/December, 121-130
- Reed, E., & Morariu, J. (2010). *State of Evaluation 2010: Evaluation Practice and Capacity in the Nonprofit Sector*. Innovation Network.
- Roberts Enterprise Development Fund (REDF) (2001). *Social Return on Investment Methodology: Analyzing the Value of Social Purpose Enterprise Within a Social Return on Investment Framework*.
- Roche, C. (1999). *Impact Assessment for Development Agencies: Learning to Value Change*. Oxfam GB, Oxford, UK.
- Rosenberg, L.J., and Posner, L.D. (1979). *The Logical Framework: A Manager's Guide to a Scientific Approach to Design and Evaluation*.
- Scott, J.C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press: New Haven.
- SIDA. (2004). *The Logical Framework Approach: A Summary of the Theory Behind the LFA Method*.
- SIDA. (2006). *Logical Framework Approach – With an Appreciative Approach*.
- Stark, D. (2009). *The Sense of Dissonance: Accounts of Worth in Economic Life*. Princeton: Princeton University Press.
- Urban Institute (2006). *Building a Common Outcome Framework to Measure Nonprofit Performance*.
http://www.urban.org/UploadedPDF/411404_Nonprofit_Performance.pdf, Accessed 19 September 2011.
- Wallace, T., Bornstein, L. & Chapman, J. (2007). *The Aid Chain: Coercion and Commitment in Development NGOs*. Practical Action Publishing: Rugby, UK.
- Waysman, M., & Savaya, R. (1997). Mixed method evaluation: a case study. *American Journal of Evaluation*, 18, 227-237.
- W. K. Kellogg Foundation. (2004). *Logic Model Development Guide: Using Logic Models to Bring Together Planning, Evaluation and Action*.
- Weber, M. (1904/1949). *The Methodology of the Social Sciences*, E. A. Shils and H. A. Finch (ed. and trans.), New York: Free Press.
- World Bank. (undated). *The LogFrame Handbook: A Logical Framework Approach to Project Cycle Management*.

World Bank (2006). *Conducting Quality Impact Evaluations under Budget, Time and Data Constraints*. Independent Evaluation Group.

Table 1: Analysis of evaluation approaches

Factor	Logical Framework	Most Significant Change	Social Return on Investment
<i>Material output(s)</i>	A 4x4 matrix. Arrayed vertically are inputs, outputs, purpose and goal. Arrayed horizontally are narrative summary, objectively verifiable indicators, means of verification, and important assumptions	Most significant change stories	Social return on investment report, which features SROI metrics along with business data, descriptions of the organization/project, case studies of participants' experiences
<i>Stated problem approach is addressing</i>	Evaluators could not compare what was planned with what actually happened, and evaluation was an adversarial process	Existing methods focused on intended outcomes using pre-defined indicators	Evaluation of social enterprises based on extent of grant making and fundraising and not on social value generated
<i>Origins</i>	Complex space age and military programs	Complex social change programs	Private sector investment analysis
<i>Purpose of evaluation</i>	To compare the outcome of the project against pre-set standards of success	To identify unexpected changes and uncover prevailing values. Avoid use of pre-set objectives and focus on making sense of events after they have happened	To use the estimated social return of an enterprise/project in evaluations of performance and/or funding decisions
<i>Preferred form of knowledge</i>	Indicators that are objective and verifiable	Stories and descriptions. Explicit aversion to indicators.	Quantifiable, financial.
<i>Preferred methods of producing knowledge</i>	Scientific methods, specification of interlinked cause-effect relations in form of 'if, then' statements, collection of pre-set indicators, usually by office staff	Story-telling and thick description by those closest to where changes are occurring. Purposive sampling.	Conduct a study with independent and objective researchers. Use statistically robust sampling procedures.
<i>Abstraction from underlying activities</i>	Project activities are translated into a hierarchy of objectives, namely inputs, outputs, purpose and goal	Activities are translated into stories that are told by those closest to the activities	Activities translated into financial terms and then converted into an index
<i>Role of external advisers</i>	None stated	Judges of 'best' stories drawn from within the organization. No explicit role for external experts	External persons can act as researchers, conduct analysis and provide analytical support
<i>Expertise</i>	Completion of 4x4 matrix. Skills in project design and specification	Technique requires no special professional skills.	Calculation of SROI metrics and indices. Knowledge of concepts such as time

	of objectives and indicators for each of the four levels of the hierarchy. Evaluators need minimal skill beyond checking how actual results compare to pre-set standards.		value of money, discounted cash flows, attribution, deadweight and displacement
<i>Conflict</i>	To be reduced/avoided through specification of clear objectives and accountabilities	To be addressed explicitly by providing a forum for surfacing and debating prevailing values	Not explicitly addressed
<i>Method of learning</i>	Deviations from pre-set standards provide opportunity to adjust programs	Focus on unexpected and exceptional events provides scope for learning	Use SROI reports to analyze how social value is generated

Table 2: Evaluation logics in the third sector

	Scientific Evaluation Logic: evaluation echoes the scientific method, with a strong focus on systematic observation, gathering of observable and measurable evidence, and a concern with objective and robust experimental procedures		Bureaucratic Evaluation Logic: evaluation is rooted in ideals of rational planning, with a strong focus on complex, step-by-step procedures, the limiting of deviations from such procedures, and analysis of the achievement of intended objectives		Learning Evaluation Logic: evaluation privileges an openness to change and the unexpected, the incorporation and consideration of a wide range of views and perspectives, and a focus on lay rather than professional expertise	
Questions	Ideal(s)	Examples from evaluation techniques	Ideal(s)	Examples from evaluation techniques	Ideal(s)	Examples from evaluation techniques
1. What makes a 'quality' evaluation?	<i>Proof:</i> claims made about effects of projects must be proved through the use of evidence. Alternative explanations must be accounted for.	<ul style="list-style-type: none"> ▪ 'indicators demonstrate results' (LFA) ▪ Account for 'attribution' and 'deadweight' (SROI) ▪ Discounting social impact (BACO) 	<i>Categorization:</i> activities and/or effects of projects must be translated into pre-defined categories	<ul style="list-style-type: none"> ▪ Project activities translated into four categories: 'inputs', 'outputs', 'purpose' and 'goal' (LFA) ▪ Value generated by projects translated into three categories: 'economic value', 'socio-economic value' and 'social value' (SROI) ▪ Financial, customer, internal processes, learning perspectives (BSC) ▪ Common outcome indicators (COF) ▪ 'Total Output' and 'Social impact' (BACO) 	<i>Richness:</i> evaluation should focus on analyzing the fullest possible range and variation of project effects	<ul style="list-style-type: none"> ▪ 'thick description' (MSC) ▪ impact of nonprofits can be 'intended or unintended, positive or negative' (IPAL) ▪ Case studies of participants' experiences (SROI) ▪ 'multidimensional framework for measuring and managing nonprofit effectiveness' (BSC)
2. What characterises a 'good'?	<i>Objective:</i> evaluative judgements and data collection processes	<ul style="list-style-type: none"> ▪ 'Objectively verifiable indicators' (LFA) 	<i>Sequential:</i> the evaluation proceeds according to a step-by-	<ul style="list-style-type: none"> ▪ The '10 stages of a NEF SROI analysis', Use of 	<i>Egalitarian:</i> evaluation approach uses techniques and	<ul style="list-style-type: none"> ▪ 'no special professional skills' (MSC)

<i>evaluation process?</i>	are not influenced by personal feelings or preferences	<ul style="list-style-type: none"> ▪ 'independent and objective researchers' (SROI) 	step process. Each stage must be completed before moving to the next stage.	'checklists' in order to proceed to next stage (SROI)	concepts that are readily understood by participants without need for training	<ul style="list-style-type: none"> ▪ Use of 'story-telling' (MSC) ▪ Use of 'change journals in which staff record the informal feedback and changes that they observe in their daily work' (IPAL) ▪ Avoid 'lengthy academic evaluations and complex, meaningless statistical analyses' (COF)
	<i>Anti-conflict:</i> avoidance of disagreements and conflicts during evaluation process	<ul style="list-style-type: none"> ▪ move away from evaluation as an 'adversarial process' (LFA) 				
3. <i>What is the focus of evaluation?</i>	<i>Reductive:</i> evaluation process aimed at representing outcomes of projects in a simplified form	<ul style="list-style-type: none"> ▪ 4x4 matrix (LFA) ▪ financial metrics and indices (SROI) ▪ scorecard (Balanced Scorecard) ▪ BACO ratio (BACO) ▪ Outcome sequence chart (COF) 	<i>Intended effects:</i> evaluation focused on analyzing whether intended effects eventuated	<ul style="list-style-type: none"> ▪ comparison of results against project objectives (LFA) 	<i>Belief revision:</i> evaluation focused on revising beliefs about the process and outcomes of projects and the evaluation technique itself	<ul style="list-style-type: none"> ▪ identify unexpected changes' (MSC) ▪ focus on 'what is working, why it is working, what must be sustained and what must be changed' (IPAL) ▪ use of outcome data 'to identify where results are going well and where not so well' (COF) ▪ 'learn more about how..input..contributes to social value creation' (SROI)
			<i>Hierarchy:</i> focus on ordering aspects of the evaluation process (e.g., activities,	<ul style="list-style-type: none"> ▪ 'hierarchy of project objectives' (LFA) ▪ 'supplemental information such as 		

			outcomes and/or data) such that some are considered higher than and/or more important than others	participant surveys' (SROI) <ul style="list-style-type: none"> ▪ overall 'mission' at the 'top' of the scorecard (BSC) ▪ 'pyramid of indicators' (IPAL) 		step 10 of MSC is 'revising the system'
4. What is the role of the evaluator?	<i>Evaluator as 'scientist':</i> evaluator conducts research and reports study findings	<ul style="list-style-type: none"> ▪ evaluator uses a 'scientific approach' (LFA) ▪ evaluator as 'researcher' (SROI) 	<i>Evaluator as 'implementer':</i> evaluator ensures that the specified evaluation process is adhered to	<ul style="list-style-type: none"> ▪ ensure activities are correctly categorized and indicators are verified (LFA) ▪ ensure stages are properly completed through use of checklists (SROI) 	<i>Evaluator as 'facilitator':</i> evaluator helps others to participate in the evaluation process	<ul style="list-style-type: none"> ▪ 'Champion' who 'facilitates selection of SC stories..encourage people...motivate people..answer questions' (MSC) ▪ Organizations encouraged to use method in a way that 'suits their needs and context' (IPAL)

Abbreviations: BACO (Best Available Charitable Option), BSC (Balanced Scorecard), COF (Common Outcome Framework), IPAL (Impact Planning, Assessment and Learning), LFA (Logical Framework), MSC (Most Significant Change), SROI (Social Return on Investment)

Figure 1: Graphic representation of 'linked hypotheses' from *The Logical Framework: A Manager's Guide to a Scientific Approach to Design and Evaluation* (Rosenberg & Posner, 1979)

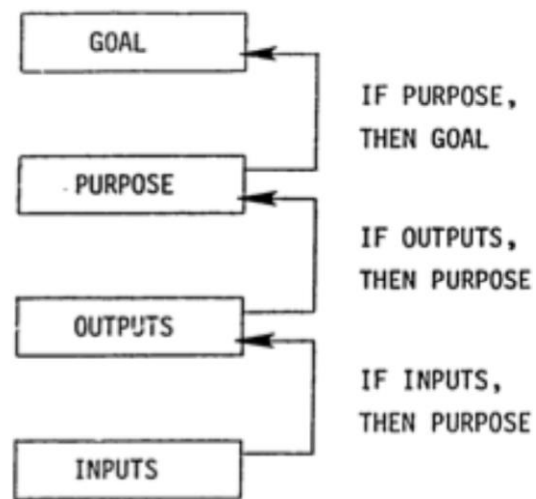


Figure 2: Image from cover page of *The Most Significant Change (MSC) Technique: A Guide To Its Use* (Dart & Davies, 2005)



Figure 3: Description of the SROI ratio from *Measuring Real Value: a DIY guide to Social Return on Investment* (NEF, 2007)

SROI measures the value of the benefits relative to the costs of achieving those benefits. It is the ratio of the net present value of the benefits to the net present value of the investment. For example, a ratio of 3:1 indicates that an investment of £1 delivers £3 in social value.

$$[\text{SROI}] = \frac{[\text{Value of Benefits}]}{[\text{Value of Investment}]}$$