On the Evaluation of Social Innovations and Social Enterprises: Recognizing and Integrating Two Solitudes in the Empirical Knowledge Base

Barbara Szijarto

Peter Milley

Kate Svensson

J. Bradley Cousins

Centre for Research on Educational and Community Services,

University of Ottawa, Canada

-- This is a draft of a paper submitted for publication consideration --

Paper presented at the annual meeting of the Canadian Evaluation Society, Vancouver, May 2017

Correspondence: Barbara Szijarto, Faculty of Education, University of Ottawa, 145 Jean-Jacques Lussier, Ottawa, ON, Canada, K1N 6N5; Tel. 613-562-5800 Email: bszijart@uottawa.ca

Vitae:

Barbara Szijarto is a doctoral candidate at the University of Ottawa. Her research interests include evaluation of complex social interventions and knowledge mobilization. She is actively involved in evaluation projects at the Centre for Research on Educational and Community Services at the University of Ottawa. She has presented nationally and internationally, most recently at the annual conferences of the American Evaluation Association, the European Evaluation Society and the Canadian Evaluation Society, as well as the Canadian Knowledge Mobilization Forum.

Peter Milley is Assistant Professor in the Faculty of Education and Senior Associate of the Centre for Research on Educational and Community Services at the University of Ottawa. Peter teaches and conducts research on topics related to leadership and policy. His recent evaluation research (with Jill A. Chouinard) includes "Mapping the spatial dimensions of participatory practice" in Evaluation and Program Planning, and "From New Public Management to New Political Governance: Implications for

evaluation" in the Canadian Journal of Program Evaluation. Peter has significant experience in leadership development, policy research, and workplace learning from a career in the Government of Canada.

Kate Svensson is a doctoral candidate at the University of Ottawa where she is pursuing her PhD in Education, with a particular focus on program evaluation. Her research interests include studying the potential uses of digital technology and social media for stakeholder learning in the field of program evaluation, and in using these tools for mobilizing evaluation knowledge.

J. Bradley Cousins is Professor of Evaluation at the Faculty of Education, University of Ottawa. Cousins' main interests in program evaluation include participatory and collaborative approaches, use, and capacity building. He received his Ph.D. in educational measurement and evaluation from the University of Toronto in 1988. Throughout his career, he has received several awards for his work and has published many articles and books on evaluation. Cousins was Editor of the Canadian Journal of Program Evaluation from 2002 to 2010 and as of 2017 is a Fellow of the Canadian Evaluation Society.

Abstract

Social innovation (SI) is billed as a new way to address complex social problems. Interest in SI has intensified rapidly in the last decade, making it an important area of practice for evaluators, but a difficult one to navigate. Learning from developments in SI and evaluation approaches applied in SI contexts is challenging because of 'fuzzy' concepts and silos of activity and knowledge within SI communities. This study presents findings from a systematic review and integration of 41 empirical studies on evaluation in SI contexts. We identify two isolated conversations: one about 'social enterprises' (SEs) and the other about non-SE 'social innovations'. These conversations diverge in key areas, including engagement with evaluation scholarship, and in the reported purposes, approaches and use of evaluation. We identified striking differences with respect to degree of interest in collaborative approaches and facilitation of evaluation use. The findings speak to trends and debates in our field, for example how evaluation might reconcile divergent information needs in multilevel, cross-sectoral collaborations and respond to fluidity and change in innovative settings. Implications for practitioners and commissioners of evaluation include how evaluation is used in different contexts and the voice of evaluators (and the evaluation profession) in these conversations.

- A knowledge divide exists in the empirical literature between studies on evaluation of Social Innovation and evaluation of Social Enterprise.
- Key differences relate to engagement with evaluation scholarship, reported purposes for evaluation, approaches to evaluation and use of evaluation.
- Factors reported to influence evaluation success were similar in both SI and SE evaluation studies. These were levels of conflict, quality of relationships, time and capacity.
- Our findings speak to debates in the field about how evaluation practice can reconcile micro and macro-level information needs and accommodate fluidity and change in innovative settings.
- Our findings support a call for greater exchange of ideas between these two domains of practice.

Keywords

Social innovation (SI) is billed as a new way to address complex social needs. Interest in SI has intensified to the point that it has been described as a "global obsession" (Rogers, 2008, n.p). Organizations in public, private, non-governmental and philanthropic sectors are committing money and other resources to stimulate the growth of SI and related ventures such as social enterprises (SEs) (e.g., Corporation for National and Community Service, 2016; European Commission, 2016; Government of Canada, 2014; J.W. McConnell Family Foundation, 2017). Consequently, this is becoming an important area of practice for evaluators.

The research literature on SI and SE is growing, including with respect to evaluation. It describes practitioners applying a variety of innovative approaches in this new terrain (Antadze & Westley, 2012; TEPSIE, 2014; Patton, McKegg & Wehipeihana, 2015; Preskill & Beer, 2012). Divides appear to exist, however, between those writing about SI versus SE. In addition, "fuzzy" conceptualizations of SI and the relationships between it and SE add to the challenge of learning from the experiences of practitioners in this new field (Cunha, Benneworth & Oliveira, 2015; Edwards-Schachter & Wallace, 2015; Svensson, Szijarto, Milley & Cousins, 2016).

We encountered these divides and conceptual challenges during a systematic review and integration of empirical studies on SI evaluation for two previous studies, where our purpose was to describe what evaluation practices look like in SI contexts, what drives those practices and how they affect SIs (Milley, Szijarto, Svensson & Cousins, 2016), as well as the implications for evaluation design (Svensson et al., 2016). Here, we take up the questions: How

do empirical reports of evaluation practices in SI and SE contexts compare? What are the implications for evaluators and those commissioning evaluation in this domain?

In this paper, we describe "two solitudes" found in our sample between studies on evaluations in SI versus SE contexts. We outline where they diverge and converge. We address some key conceptual issues, and conclude by discussing how expectations about the role of evaluation in SI and SE affect the conduct and use of evaluation.

Defining Social Innovation and Social Enterprise

"Social innovation" and "social enterprise" have been defined in a variety of ways (e.g., Cunha et al., 2015; Pol & Ville, 2009; Westley, 2013). Some definitions are broad or vague and make it difficult to distinguish SI and SE from conventional social interventions (Svensson et al., 2016). Contradictions among definitions and the interchangeable use of related terms, such as the conflation of social innovation and social entrepreneurship (Westley & Antadze, 2010), likely relate to the relative youth, rapid uptake, and appeal of SI and SE across a range of fields and sectors. As part of our research (Svensson et al., 2016), we compared definitions of SI found in the conceptual literature to those in our sample of empirical studies on evaluations in SI contexts. We were looking for features that distinguish SI from other types of social intervention in ways important to evaluation. We briefly note the findings from this study below, and follow with an outline of features of SE derived through a similar review of the literature on this concept.

Social Innovation

Westley and Antadze (2010) define SI as "a complex process of introducing new products, processes or programs that profoundly change the basic routines, resources and authority flows, or beliefs of the social system in which the innovation occurs" (p. 2). The process focus has been called the "sine qua non" of SI (Hubert, 2010, p. 42). It is likely to include a heavy emphasis on collaboration and co-creation through experimentation, often explicitly drawing on "design thinking"' (e.g., Edwards-Schachter & Wallace, 2015). Collaboration among diverse actors is promoted as a way to foster creativity (e.g., Hubert, 2010) and can involve tapping into multiple funding sources, including from the private sector. The commitment to participatory approaches in SI can extend to how intended outcomes are defined, including co-defining what change is desired as part of the process of SI and resisting specification in advance (e.g., Cunha et al., 2015; Murray, Caulier-Grice & Mulgan, 2010). Language around social change, system change and transformation is common, especially with respect to vertical scaling (through levels in a system) as a complement to lateral scaling (expansion to other geographic areas or to more people) (see e.g., Westley et al., 2014).

Complexity thinking is evident in the conceptual and empirical literature on SI, emphasizing, for example, nonlinear relationships between inputs and effects, the importance of context and agency, and the emergent character of outcomes. Complexity is seen to feature in the processes inside SI initiatives (for example, social dynamics of collaboration), the problems SI targets (for example, poverty as a complex issue), and the surrounding environment (for example, rapid demographic change) (see also Preskill et al., 2014, on sources of complexity). Of note, definitions of SI sometimes assume uncritically that system changes or

transformations will be beneficial (Pol & Ville, 2009), squaring with Dahler-Larsen's (2016) observation that the term "innovation" tends to have positive connotations, even though something new is not always better.

Given the diversity among SIs and conventional social interventions, we see differences between them to be a matter of degree. The features discussed above are meant to be used heuristically as sensitizing categories to guide understanding of SI and inform evaluation design (Svensson et al., 2016), rather than as definitions in an operational sense (Patton, 2007).

Social Enterprise

Social enterprise is a term frequently associated with SI but, unlike SI, it refers primarily to characteristics of an organization rather than a process or intervention (see e.g., Antadze & Westley, 2010, Cunha et al., 2015). Although descriptions of SE vary in the literature we reviewed (see below), five recurring themes stand out.

Hybrid aims. Hybrid mixes of economic and social aims distinguish SEs from conventional not-for-profit and social sector organizations as well as from purely for-profit businesses. Although SEs are typically active in markets and aim to generate economic profits, some or all of these profits are reinvested in pursuit of their social missions. What is seen to distinguish SEs from conventional private sector organizations is their emphasis on profits as the means to achieve social ends (e.g., Arena, Azzone & Bengo, 2015; Thompson et al. 2000).

Hybrid financing. SEs are not averse to accessing multiple sources of funding, including governments, philanthropic foundations and private investors, however self-sufficiency and

non-reliance on public funds are frequently noted as goals of SE actors (e.g., Luke, Barraket & Eversole, 2013; Nicholls, 2008).

Use of business methods. SEs are seen to be more likely than conventional social service organizations to apply methods adopted from the private sector, for instance by using marketing strategies to raise demand for their goods or services.

Innovation identity. SE actors and advocates describe SEs as seeking to provide innovative responses to complex social needs left unaddressed by conventional private and public sector organizations (i.e., "filling institutional voids" in Nicholls, 2009, p. 759; see also Luke et al., 2013; Sadownik, 2012), often at a local level and/or with an interest in scaling out laterally. They also claim SEs are more nimble, flexible, efficient, transparent and accountable (e.g., Nicholls, 2009) than conventional social sector organizations. Methods akin to design thinking that are "characterized by trial and error, continuous iteration, and a focus on results" are also claimed to set SEs apart from other entities (Galvin & Iannotti, 2014, pp. 424-5). The level of importance attributed to individual "change agents" or "social entrepreneurs" in some of the discourse (Galvin & Iannotti, 2014; Vo, Christie & Rohanna, 2016) also sets it apart.

Implications for Evaluation

The foregoing distinguishing features have implications for evaluators working in SI and SE contexts. For example, the purposeful engagement of a diversity of stakeholders across sectors and social boundaries, as well as the intensity of collaboration, are likely to be greater than in more established, conventional social programs, generating tensions and conflicts of a different degree and kind. These dynamics complicate the evaluation process (Milley et al.,

2016). SE actors, for example, are described in empirical studies as being caught between the public and private spheres, "negotiating competing legitimacies" represented by diverse stakeholder demands in evaluation practice (Barraket & Yousefpour, 2013, p. 455; see also Ebrahim & Rangan, 2014; Galvin & Iannotti, 2015). Establishing a coherent conceptualization of the intervention among diverse actors is also described as a key activity in some SI evaluations (e.g., Moore & Cady, 2015). While this is generally thought to be good evaluation practice, it may be particularly poignant in the context of SE's and SI's given their unique character. We expand on this and other implications for evaluation in the discussion section of this paper.

Methods

As signaled in the introduction, our original research project from which this paper emanates aimed to map the empirical landscape of evaluation in SI contexts. The SI and evaluation fields both draw on a broad range of ideas and traditions within and across disciplinary and geographic boundaries (Milley et al., 2016; Westley, 2013). As such, we faced a choice of selecting from among available non-overlapping definitions to focus our review in a specific area, or adopting a broad view at risk of including activities in the fuzzy margins of evaluation and SI. We chose to take a broad view to correspond with the study's exploratory mapping intent. This resulted in a sample of studies that included evaluation of both SIs and SEs; the comparison of these subgroups is the focus of this paper.

Inclusion Criteria

Our initial search included any empirical articles explicitly using terms related to "social innovation", including "social enterprise".[1] In addition, we defined evaluation broadly as: "systematic inquiry to serve a range of policy and program purposes, such as enabling learning, development, improvement and capacity building, informing judgements about the merit, worth and significance of policies and programs, and supporting oversight, accountability and compliance" (Milley et al., 2016, p. 4). We restricted the search to peer reviewed, empirical studies, excluding "gray" literature and articles that did not discuss SI and/or evaluation within our broad definitions.

Search

We searched multiple databases and evaluation journals for English-language studies published between January 2000 and October 2015. To obtain an exhaustive sample, we also searched the reference lists of included articles and sought input from other researchers. After eliminating duplicates, we reviewed the abstracts or full texts of 84 studies. Of these, 43 were then excluded for the reasons noted above. This left a final sample of 41 studies.

Data Extraction, Coding and Analysis

We summarized the 41 articles using a template based on our research questions. Data were extracted by two team members working independently, following a predefined coding framework related to drivers, activities and consequences of evaluation practices. Additional

---

[1] The search string used: ('social innovation' OR 'innovation hub' OR 'social entrepreneur' OR 'social enterprise' OR 'social finance' OR 'social impact') AND (evaluation).

codes were derived inductively and discussed at intervals by the entire team to arrive at a final framework. This framework was then applied recursively to the article summaries. Team members met regularly to compare results and reconcile differences. Categorical data were analyzed to gain an overview of the sample and we used NodeXL[2] to map relationships among the studies drawing on metadata (e.g., the reference lists of articles). This led us to identify two subgroups of studies that differed on key characteristics: one group that describes evaluation in SEs (13 articles), and another that described non-SE social innovations (27 articles).[3] We then pursued in-depth qualitative analysis of the two groups separately. Themes were derived inductively by individual team members, bearing in mind our research questions. Emerging themes were developed and refined in team meetings.

Findings: "Two Solitudes" of Social Innovation and Social Enterprise

In what follows, we describe interesting points of difference and convergence between the two groups of studies, identified through our analysis of the sample (above). These include the academic or professional affiliations of authors, the study designs and knowledge bases referenced in the studies, the reported purposes for and approaches to evaluations, use of evaluative work by SIs and SEs, and factors reported to influence evaluation success.

Author affiliations

Most of the studies in the SI group were authored by individuals directly involved in evaluation work described in the study. Many were co-authored by individuals located in the

---

[2] Social Media Research Foundation: Nodexl.codeplex.com; smrfoundation.org
[3] One additional article described SE as well as non-SE evaluation.

initiative or with a funder (e.g., Allen, Hunsicer, Kjaer, Krimmel & Plotkin et al., 2015; Tan, McGill, Tanner, Carlson & Rebok et al., 2014). The involvement of practice-based authors adds valuable perspectives, such as deep contextual knowledge, and lends greater credibility than would be the case were these narratives authored by evaluators only; however, these contributions might also be seen as less impartial. Data sources and analysis steps were rarely documented in the SI studies, making it more difficult to trace the process leading to conclusions. Although some of the authors in the SE group also reported on evaluations in which they were directly involved (e.g., Jackson & Tarsilla, 2013; Pollard & Cook, 2013), all but one of the SE authors report their affiliation to be with academic institutions. SE articles tend to include explicit, formal methods descriptions. In sum, the studies in the SE group are, in general, more formal, further removed from the field of practice, and suggest a more academic perspective than those in the SI group.

Study Designs

All the SI studies are case-based; most are case narratives (78%) or single case studies (19%)[4] and only one is a multi-case study. In-depth single case narratives provide opportunity to describe evaluation activities and results in detail; studies in this sample were especially rich in their reflective descriptions of how evaluations unfolded over time (see below). In contrast, most of the SE group are multi-case studies or other research designs (e.g., qualitative interview studies); only two are case narratives (see Figure 1). Here, we see a strength in the diversity of research designs and availability of more robust designs in the SE set.

---

[4] We distinguished 'case studies' from 'case narratives' based on the methods described by the authors.

< Figure 1 here. >

Knowledge bases

We were surprised by differences in use of program evaluation literature between the two sets. The authors in the SI group drew heavily on evaluation literature. Nearly all (93%) referenced evaluation sources and the majority (86%) referenced evaluation scholarship.[5] In contrast, only five articles (38%) in the SE group included scholarly sources from the evaluation literature and nearly half (46%) made no reference to evaluation sources of any kind (including gray literature).

We mapped the studies' references using social network analysis software (see Figure 2). We identified almost no common sources between the SE and SI groups. Of 1268 items in the combined reference lists; only 2 sources were common between the groups, both of which were research methods texts. Within the SI group, some commonality existed in source material, in particular, work on utilization-focused and developmental evaluation by Patton (1994, 2008, 2011), Preskill and Beer (2012) and Gamble (2006, 2008). Westley, Zimmerman and Patton's (2006) work on SI was also cited numerous times in the SI group. Within the SE group, we see no evidence of a shared knowledge base of this kind[6].

Finally, most of the SI studies were published in program evaluation journals or an edited book on evaluation; none of the SE studies were found in program evaluation

---

[5] 'Scholarly' evaluation sources were defined as any published in peer-reviewed program evaluation journals, or textbooks on evaluation. 'Gray' sources included non-peer reviewed materials, e.g., online evaluation toolkits.
[6] The most-cited text in the SE group was Dart (2004), by 4 of the 13 studies.

publications. Overall, these patterns of engagement with evaluation scholarship are

remarkable, and we offer possible explanations for them in the discussion.

< Figure 2 here >

Evaluation Purpose

Studies in both groups tended to describe multiple drivers of evaluation. Most of the SE

studies (62%) identify accountability as the primary purpose, followed by program

improvement and tactical/political purposes. An opposite pattern exists in the SI group where

program development was identified most often as the primary purpose (70%), followed by

program improvement. Accountability purposes were reported to be important in SI evaluation,

but rarely as a primary driver; we did not see political purpose reported as a primary driver (see

Figure 3). In distinguishing between improvement and development purposes, we followed

Patton's (2011) description of improvement in terms of incremental adjustments or "fine-

tuning" to make a model or intervention better, versus development, which involves

substantial changes to the initiative, such as clarifying the core purpose of the initiative or

redesigning services (pp. 39-40).

< Figure 3 here >

The multiple purposes for evaluation (both primary and secondary) that were reported

by these authors can be grouped loosely in two categories: external accountability and

legitimacy needs, and internal decision-making and influencing needs. External accountability

and legitimacy needs were dominant among the SE studies and included meeting funder

reporting requirements, new resource acquisition, and demonstrating mission alignment with

external stakeholders. These needs align with comparative, standardized metrics or models that aim to increase transparency and/or allow external stakeholders to compare across SEs to inform funding or investment decisions. Among the SE studies, the second category (internal decision-making and influencing needs) was smaller and includes both monitoring and evaluation (M&E) aimed at organizational-level performance improvement, identity-development and internal legitimacy, momentum and morale. Tailored, context-attentive M&E methods align better with this second set of needs.

Most authors in the SI set referred to interest in learning or development of the initiative as directing evaluation. A majority also referred to the complexity of the initiatives influencing chosen approaches. Only two studies in the SI group were primarily concerned with summative/impact evaluation (Saari & Kallio, 2011; Tan et al., 2014), both were described as collaborative, and one was described as developmental in intent, with the aim to enhance learning among participants through understanding impact. Decisions about evaluation approach were largely centred in the initiative, in contrast to the SE group where decisions were largely focused at the organizational level or on its external stakeholders.

Approaches to Evaluation

The diversity of approaches described in the SE set was high, but precision in the use of terms and concepts was sometimes low, creating challenges for analysis. For example, although Galvin and Iannotti (2015) differentiated between the terms evaluation and monitoring, and Jackson and Tarsilla (2012) distinguished between evaluation and social accounting, other SE authors used evaluation as an umbrella term (e.g., Arasti, Zarei, & Didehvar, 2015), or used

evaluation, performance measurement, impact assessment or other terms interchangeably
(e.g., Palmas & Lindberg, 2013).

Correspondingly, SE authors discussed more than a dozen frameworks, methods or
approaches across a spectrum of evaluation-related endeavors. Most can be grouped under
impact assessment or performance measurement categories (see Figure 4). They included
frameworks drawn from social accounting, including Balanced Scorecard and Social Return on
Investment (SROI) (see e.g., Arena, 2015; Luke et al., 2013; Nicholls, 2009), as well as general
methods such as mixed-method case comparisons, logic model analysis (e.g., Jackson & Tarsilla,
2014), and pre-post outcomes measurement (Galvin & Iannotti, 2014). Three articles did not
describe approaches or methods in sufficient detail to support this analysis (e.g., Barraket &
Yousefpour, 2013; Palmas & Lindberg, 2013).

Explicit reference to the field of program evaluation (i.e., as developed for the social
services sector) was sometimes made in the SE studies, but lightly. This reinforces our
impression that the links between the SE sector activities and the field of program evaluation is
tenuous, at least as described in this literature. The exceptions Jackson and Tarsilla (2014), who
pointed to specific frameworks (e.g., Contribution Analysis) and suggested SE actors look to
"the field of program evaluation [as] an important source of methodological experience and
tools" (p. 135); and Sadownik (2012) who mentioned critique of impact measurement in the
program evaluation literature.

In contrast, SI authors focused on approaches from the field of program evaluation, and
all described evaluation activities in detail. Discussion of social accounting frameworks,

prevalent in the SE sample, were not found here. A majority described the use of

developmental evaluation (DE) approaches or adaptations thereof (Patton, 2011). Authors

reported the use of a range of methods and techniques, including some used in conjunction

with a DE approach for outcomes measurement to support summative decision-making (e.g.,

Outcome Harvesting, see Wilson-Grau et al., 2015; Contribution Analysis, see Cabaj et al.,

2015). Extensive use of mixed methods was reported, as well as affirmative methods (e.g., Most

Significant Change, Mathie & Peters, 2014). Nonetheless, we questioned why more diversity,

such as the use of realist or experimental designs was not found, especially when there are

indications that more diversity exists in practice (e.g., Preskill and Beer, 2012; TEPSIE, 2014).

This may be due to the dominance of DE in our sample (discussed below).

< Figure 4 here >

Collaborative practices. We used magnitude coding[7] to categorize the studies based on

reports of collaborative practices (see Figure 5). Our codes drew from the 8 evidence-based

principles of collaborative approaches to evaluation, developed and validated by Shulha et al.

(2016). A majority of the SI studies focused on collaboration as a key theme and dedicated

substantial space to description of collaborative activities across multiple phases of evaluation.

In contrast, those SE studies that did report sufficient detail about activities suggested less

collaboration in the evaluative work.

< Figure 5 here >

---

[7] 'Magnitude Coding', i.e., Low (0-2 principles reported to have been applied), Mid (3-5), High (6-8). NA or NR indicates the study did not report sufficient detail in the empirical findings to assess or the study design did not generate this type of information.

Evaluation Use and Influence

We also reviewed the findings of these 41 studies for reports about the use and influence of evaluation in SI and SE contexts. The authors of SI studies generally reported a spectrum of uses, including conceptual, instrumental and tactical uses (Weiss, 1979, 1998). For example, Allen et al. (2015) reported the evaluation influencing how the initiative envisioned success (conceptual use), leading to improvements in a grant solicitation process (instrumental use). Most studies also referred to use stemming from the evaluation process (process use) as well as from evaluation findings (Patton, 2007). As an example of process use, Asher et al. (2015) described the evaluation process helping actors "carve out" time for regular long-term strategic thinking which otherwise competes with urgent day to day work; this thinking was integral to the "reframing" of the initiative, which led to changes in structure and resourcing (pp. 123-4).

Nearly all the SI studies described the influence of evaluation activities on the SI as beneficial: for example, advancing vision or strategy of the SI; contributing to the improvement of the SI's processes; helping stakeholders address differing information needs at aggregate and local levels; increasing the capacity to do and use evaluation (for this distinction see Cousins et al., 2014); helping to reduce risks associated with SI implementation (for more detail, see Milley et al., 2016).

In contrast, the authors of SE studies were mostly silent about use; only a small number reported on use in their empirical findings. These authors reported use in terms of instrumental or tactical use of evaluative findings. Examples of instrumental use were rare; an example is

evaluation findings contributing to a decision to renew a policy (Jackson & Tarsilla, 2012).

Examples of political or tactical use were somewhat more prevalent, including use for funding

acquisition, or increased legitimacy (Luke et al., 2013; Nichols, 2009). Barraket and Yousefpour

(2013) reported limited evidence of a direct relationship between the evaluative activities and

organizational operations, beyond leveraging new funding and communicating success for

internal morale.

Almost one-half of the SE articles (46%) made no mention of overall influence of

evaluative work on SEs in their empirical findings. Among those that did, negative effects stand

out more than positive ones. Four studies noted a high, sometimes unwarranted, resource

burden placed on SEs by M&E activities (i.e., Arena et al., 2015; Barraket & Yousefpour, 2013;

Luke et al., 2013; Ndemo, 2006). Descriptions of M&E activities contributing to goal

displacement or mission drift appeared in four studies (Hazenberg, Seddon & Denny, 2014;

Luke et al., 2013; Nicholls, 2009; Palmas & Lindberg, 2013). Although Arena et al. (2015)

advocated for more performance measurement on the basis that lack of performance data

discourages investors, other authors suggested some approaches to M&E may disadvantage SEs

relative to purely for-profit competitors (Hazenberg et al., 2014; Luke et al., 2013; Nicholls,

2009; Palmas & Lindberg, 2013). For example, Hazenberg et al. (2014) described outcome data

comparing an SE to a competitor that obscured the SE's "willingness to take on more 'socially

excluded' individuals" (p. 897). While there is a strong need to demonstrate achievement, some

cautioned that social outcomes may be devalued if they are more difficult to measure or

interpret. Nicholls (2009) wrote that in the context of social accounting: "many successful social

enterprises, therefore, appear to be barely profitable and, consequently, risky contractual partners compared with other private sector companies" (p. 766).

Factors Influencing Evaluation Success

We identified factors that appear to mediate the effects of evaluations on SIs and SEs. We saw the most similarity between the two groups of studies in this area. These included levels of conflict, quality of relationships, time and capacity (Milley et al., 2016).

Conflict and relationships. Across the SI studies, authors identified conflicts within and between various groups of actors who have different ideas about the initiatives and/or the evaluation of the initiatives. The emphasis on multi-sectoral collaboration, along with the experimental and uncertain character of SIs, may mean conflict is an inherent feature of SI (see Cabaj et al., 2015; Tan et al., 2014; Togni et al., 2015). We also noted that the emergent and dynamic nature of evaluation approaches such as DE or empowerment-oriented evaluation (see Mathie & Peters, 2014) may be stressful for actors accustomed to more highly structured processes (see e.g., Murphy, 2015; Togni et al., 2015) and for evaluators navigating hybrid internal/external roles (e.g., Poth, Pinto & Howery, 2012; Rey, Trembley & Brouselle, 2014). SI authors, particularly in the DE studies, reported placing emphasis on, and some success in overcoming conflict, and developing stronger relationships among actors over the course of evaluation, including addressing tension between macro- and micro-level information needs (e.g., between summative decision-making needs and needs for ongoing development of local initiatives) (e.g., Cabaj et al., 2015; McKegg et al., 2015; Wilson-Grau et al., 2015). Wehipeihana et al. (2015) described relationships as "the anchor providing balance and reassurance in times

of uncertainty" (p. 42). Authors offer ideas about how evaluators can build relationships,

emphasizing strong facilitation skills (e.g., Wehipeihana et al., 2015), clear communications

(Moore & Cady, 2015; Tan et al., 2014), the use of appreciative and participatory methods (e.g.,

Langlois, Blanchet-Cohen & Beer, 2013; Mathie & Peters, 2014; Moore and Cady, 2015; Zamir

and Abu Jaber, 2015) and reflective practice sessions to create "neutral" spaces places to build

trust and understanding through dialogue and critical thinking (e.g., Gamble et al., 2015;

Howard et al., 2015; Togni et al., 2015; Moore and Cady, 2015).

In the SE group of studies, authors also reported conflicts and tensions stemming from

competing interests among actors and stakeholders around measurement practice (see Galvin

& Iannotti, 2015; Hazenberg et al., 2014; Luke et al., 2013; Ndemo, 2006; Nicholls, 2009; Palmas

& Lindberg, 2013). Misalignment or lack of deliberation about information needs at different

levels and for different audiences, and how to generate information through evaluative work,

featured in eight studies. For example, Galvin & Iannotti (2015) described a case in which

funder pressure for an RCT evaluation was resisted by the SE, because of concern that real-

world implementation made a valid control group impossible and could obscure the SE's

impact. Arena and Azzone (2015) described very different information priorities among

stakeholder groups contributing to conflict and creating challenges for determining how to

operationalize "social impact". A few of the SE studies also discussed effects of weak

relationships. For example, Hazenberg et al. (2014) reported lack of trust and poor quality

communication between SE and funder contributing to emphasis on outcome targets, in turn

placing pressure on the SE to alter operations in a direction that misaligned with mission. These

authors advocated for improving relationships between SEs and funders so that they can move

beyond financial targets to a richer discussion of what it means to achieve a social mission. Yu and McLaughlin (2013) directed readers to research demonstrating that evaluation is "most effective when integrated into the organization's day-to-day operations and culture on the ground" (p. 34) and advocated for better communications between front-line staff and higher level decision makers, and greater involvement of staff and volunteers in evaluation. Arena et al. (2015) talked about advantages of collaborative development of indicators and present this as a component of a new model, intended to address what they describe as a gap in the field.

Time and capacity. Most of the SI studies pointed to the importance of long-term engagement with an initiative for evaluation to have a positive influence. Several studies described multi-year evaluations enabling the evaluators to understand context and information needs, develop trust with stakeholders, and adapt the evaluation design to the needs of different groups as these needs became better understood (e.g., Asher et al., 2015; Moore & Cady, 2015; Langlois et al., 2013). For example, Cabaj et al. (2015) described a traditional evaluation that was shifted to a developmental approach because conflict over methods led to risk of some stakeholders disengaging. The authors attributed participation by the funder in collective learning events to the development of trust and support for changes to the evaluation design; they stated that strong relationships and a sense of shared ownership were essential to the success of the evaluation. They noted that new staff arriving later in the process "were substantially less 'ready, willing, and able'" to work with the evaluation process (p. 190). Capacity building was reported as an intentional component of the evaluation in eight of the 27 studies. For example, Gamble et al., (2015) describe staff taking responsibility for

evaluation activities with the evaluator in a role of coach, in order to build internal capacity (see

also e.g. Cabaj et al., 2015; McKegg et al, 2015; Rey et al., 2014).

Time is also mentioned by two of the SE studies with respect to the time it takes for high

quality measurement, for example, to design data collection instruments, recruit participants

and do analysis (e.g., Jackson & Tarsilla, 2012). A more prevalent theme (raised in 6 studies) is a

lack of technical skills and capacity within the SE acting as a barrier to M&E activities. For

example, Arasti et al. (2015) describe insufficient evaluation capacity building, limited

stakeholder participation and unrealistic expectations as detriments to evaluation (see also

Luke et al., 2013; Sadownik, 2012). Sadownik also suggests capacity may be an issue among

funders, in the sense that many investors require reporting of "impact" "without being clear

about what they are asking for, and generally offering little guidance or support on how an

applicant or investee might provide this information" (p. 159).

This review included a rich and diverse sample of 41 studies, assembled through

intentionally inclusive search criteria and broad definitions of SI (including SE) and evaluation.

These studies presented a view of active experimentation and innovation in evaluation and

related measurement practices, as efforts to meet the needs of actors who are themselves

engaging in novel processes and structures through SIs and SEs. The SI studies described cases

in which evaluators experimented with flexible evaluation designs, prospective forms of

evaluation, systems thinking, hybridized team structures, collaborative approaches and creative

boundary spanning in response to conflict and tensions. The summative examples in this set

also emphasized collaboration and adaptability in design to accommodate stakeholder needs

and the implementation context (e.g., Tan et al., 2014, see also Cabaj et al., 2015). Some of the

SE studies presented new or adapted models or frameworks for M&E tailored to the SE context

(e.g., Arena et al., 2015; Sadownik, 2012), and several discussed opportunities for

methodological advances in social accounting, performance measurement and evaluation to

better serve the unique needs of SEs located in-between the for-profit and non-profit sectors.

Both groups of studies provided examples of innovation-in-action, suggesting there is no one

best way to evaluate SIs or SEs. However, the experimentation, and presumably learning, in the

two domains appears to be taking place in isolation from each other. Joint conversation would

be fruitful. We close this paper with discussion of a few key issues for dialogue at the level of

practice, the field of evaluation, and research on evaluation.

Discussion

Evaluation Purposes, Approaches and Uses

There are a few likely explanations for the emphasis on impact and accountability in the

SE group and the emphasis on development and learning in the SI group. We note the influence

of DE literature in the SI sample, an approach which aligns itself with complexity perspectives

(also seen in the broader SI literature), and a concomitant focus on process and adaptive

learning[8]. There was no similar common knowledge base in the SE group, suggesting a more

diverse and isolated set of voices, and a set drawing minimally from evaluation theory and

thinking, DE or otherwise. The influence of social accounting in the SE group, not found in the SI

---

[8] The proportion of DE studies in the scholarly literature may over-emphasize a learning focus in SI
evaluation practice in general: We note work in the SI 'gray literature' suggesting accountability
purposes are prevalent (i.e. Corporation for National and Community Service, n.d.; Preskill & Beer, 2012;
TEPSIE, 2014)**.** In like fashion, some of the SE authors describe evaluation practice in the SE domain as
descriptive and anecdotal, even while their own studies centre on formal M&E methodologies.

studies, is linked to a stronger relationship with the for-profit sector, and may also relate to how SEs are conceptualized. SEs are essentially organizations, with more measurement focus at the organizational level (organizational performance, impact of the organization), unlike SIs that are conceived of as processes or interventions led by collaborations.

We identified striking differences with respect to emphasis on facilitating evaluation use. As noted above, the SI studies describe evaluation designed mostly to meet internal needs, and most report use of evaluation to advance the initiative. In contrast, the SE studies described largely externally-focused M&E, relatively little discussion of evaluation use as a topic outside political or tactical use, and a number of negative effects. There are a few possible explanations for this as well. A relationship likely exists between the different approaches described and observations about use. It is not too surprising, for example, to see reports of conceptual and process use in highly collaborative and participatory evaluations designed to emphasize these forms of use. Lower reported levels of political and tactical use are consistent with participatory evaluation research (Cousins & Chouinard, 2012). Second, the robust reports about evaluation use in the SI group may stem from the authors' roots in the field of program evaluation which has a long history of interest in use and an established "use theory" (e.g., Cousins & Leithwood, 1986; King, 1988; Patton 2007; Shulha & Cousins, 1997; Weiss, 1988, 1997). In contrast, based on the sources cited in their studies, most authors in the SE group do not appear to be aware of, or perhaps see value in, the program evaluation literature. To us (positioned as we are in the program evaluation field), it is surprising to see research on evaluation taking place in SE contexts without reference to four or more decades of program evaluation literature, even more so when claims are made that the academic literature on

"performance in the social sector" is limited (it "lags behind", concepts are under-theorized) (Ebrahim & Rangan, 2014, p. 119). For example, Cordes (2016) has recommended that the evaluation field's experience implementing cost benefit analysis with social programs be applied to development of SROI. Many of the issues raised by authors in the SE group in their critiques of M&E practice (e.g., misalignment of information and needs, poor communication, goal displacement, evaluation limited to political/tactical uses) are researched and discussed extensively in the program evaluation literature, and a variety of evaluation approaches (e.g., utilization-focused, participatory) have been developed to address them. A third possible explanation for differences in how authors describe use relates to the study designs in the two groups. As noted above, the SI studies tended to be less formal single-case reflective narratives and case studies, while the SE group was comprised of a greater mix of study designs and tended to more formal methodologies, including what we perceive to be a more critically reflective stance vis-á-vis measurement in SE. We discuss this further in implications for research below.

Finally, we should note that just as we are reminded that "innovation" should not always be assumed to be a good thing (Dahler-Larsen, 2016), we should not assume that "use" is automatically positive. There will be differences of perspective and in accepted criteria about what information counts as high quality or "valid" or "rigorous" or "meaningful" enough to warrant use. This is not a question that this review is designed to address, however we suggest that communication, capacity, collaboration, time, trust - in effect the mediating factors discussed above - are essential precursors to *informed* methods and *informed* use, and likely to help limit mistaken use and misuse of evaluation (Cousins, 2004).

Trends and Debates

Reconciling multiple levels and "logics". The tension between internal and external

information needs was a dominant theme in this review. In the SE group, tailoring M&E to

individual SEs needs was described as unsatisfactory because it limits comparison across SEs

that can generate knowledge to advance the sector (Arena et al., 2015, Yu & McLaughlin 2013).

Arena et al. (2015) point to international bodies (e.g. the European Commission, social

investment funds) seeking aggregate measures of impact that can inform the larger "social

impact investing system" (p. 669). Yet authors also caution that the interest in comparison

across SEs can have unanticipated consequences. Their criticisms center on issues of reliability

and validity, pointing out that assigning value to some social benefits is more difficult than for

others (e.g., employment versus social inclusion) and how the degree of "assumption and

approximation" involved can make such values misleading (Luke et al., 2013, citing Gibbon &

Dey, 2011, p. 239; see also Hazenberg et al., 2014; Nicholls, 2009). They also argued that

metrics are likely to vary according to stakeholder perceptions of value (Luke et al., 2013) and

that universal metrics are insensitive to cultural differences in understanding "social value"

(Lane & Casile, 2011, cited in Barraket & Yousefpour, 2013; see Chouinard, 2016; Chouinard &

Hopson, 2015 on cultural sensitivity and evaluation). Some also noted there is a risk of goal

displacement toward indicators valued by funders (Luke et al., 2013; Palmas & Lindberg, 2013),

with one study reporting the potential of reduced diversity in the SE ecosystem in a region

(Palmas & Lindberg, 2013) because exclusive interest on a single metric in a funder's evaluation

framework narrowed the range of initiatives "worth" pursuing.

Nicholls (2009) characterized the field of M&E for SEs as "emergent" (p. 756) and observed two competing dynamics. On the one hand, he saw innovative measurement practices (resembling 'bricolage') being used to address multiple "logics" and needs associated with the hybridized organizational forms, missions and diverse stakeholder groups of SEs. On the other hand, the adaptation of accounting methods to the SE domain may be influenced by preferences of certain stakeholders and desire for standardization rather than information needs of SEs. Nicholls also saw a trend towards the adoption of an "investment" orientation to funding by philanthropists, foundations and governments leading to a narrowing of acceptable approaches to measurement, especially those associated with internal process needs (learning, effectiveness, innovation) (p. 766). Vo et al. (2016) pointed to published definitions of impact measurement and social accounting, which reflect a trend over time to "emphasize non-monetized outcomes and represent broader stakeholder perspectives", yet are incongruent with what they saw as a persistent overall emphasis on accountability driving outcomes measurement practice and which they attribute to demands of "funders, taxpayers, clients and concerned citizens" (p. 472).

Authors also raised concerns about measuring impacts at a system or population level (e.g., reduced prevalence of disease), which are difficult to demonstrate or attribute to an individual SE's activities (see Galvin & Iannotti, 2015; Sadownik, 2012). Authors noted that system impacts are often very long term and might be different than predicted in advance (Yu & McLaughlin, 2013). This is consistent with Ebrahim & Rangan's (2014) argument for improving understanding of the differences between program level, organizational level and system or societal level purposes and goals of evaluation. The emphasis may be changing

among some actors. Sadownik (2012) noted that some foundations are moving away from "the

constraint of objectively demonstrating the impact of a specific grant" to approaches that

consider a range of activities and provide information that is current and actionable (p. 145).

Sadownik pointed to supporting research from program evaluation that describes interventions

as complex and which suggests focus on impact measurement may interfere with learning,

collaboration and innovation. Examples in both groups of studies suggest that it is possible to

successfully respond to needs at different levels. Jackson and Tarsilla's SE case study highlighted

targeting micro level (e.g., client outcomes), meso-level (e.g., patterns of client use of the

service), and macro level (e.g., unemployment rate). Cabaj et al.'s (2015) SI case study similarly

provided an example of multi-level evaluation to meet different stakeholder needs, with

learning-focused work for pilot project stakeholders, combined with gathering evidence of

effectiveness for summative decisions. Both studies emphasized capacity, time and strong

relationships for making this possible.

Authors in the SI group offered some critique of evaluation activities in relationship to

the wider SI domain, though this was generally more muted than their SE counterparts. For

example, Mathie & Peters (2014) wrote that funder demands for accountability to specific

results may have good intentions, but they produce a "dance of deception"[9] that results in

evaluation as "window dressing" in which: "Evaluation becomes more about convincing the

source of funding that the desired results have been achieved, and less about informing course-

of-action decisions at strategic intervals" (Mathie & Peters, 2014, p. 408, see also Saari & Kallio,

---

[9] Mathie and Peters (2014) attribute the phrase 'dance of deception' to Tim Brodhead, McConnell
Family Foundation.

2011). In a case narrative, Murphy (2015) described how members of the initiative: "had all experienced collecting and reporting standardized data to satisfy funding requirements, despite the fact that the data collected were not always useful, were sometimes burdensome to obtain, and often oversimplified the complexity of the phenomena being evaluated" (p. 66).

Fluidity and change. Authors also pointed to traditional evaluation methods which tend to emphasize prespecified outcomes and assume program logic can remain stable over time, hence emphasis on predefined targets and importing "proven" models. A role has been suggested for evaluation to help balance necessary adaptations of an innovative initiative with fidelity to "key ingredients" (Dufour, Lessard & Chamberland, 2014); others cautioned evaluators from "imposing order on situations that may remain fluid and changing" (Lam & Shulha, 2015, p. 361), for example by prescribing indicators in advance; SI may require thinking in different ways (Lam & Shulha, 2015, p. 360). Dickson & Saunders (2014) echoed this, mentioning "limitations of standard evaluation practices" for programs operating in complex and changing environments (p. 177; see also e.g., Howard et al., 2015; Langlois et al., 2013). Some argued that standard methods are likely to inhibit innovation, whereas "evaluation methods that can shed light on failure, and even embrace it, are more likely to provide the insight required to refine the innovation or change course as necessary" (Mathie & Peters, 2014, p. 408). Authors urged complexity-informed evaluation methods as better suited to SI (e.g., Dickson & Saunders, 2014; Lam & Shulha, 2015; Mathie & Peters, 2014). Dickson and Saunders (2014) suggested that evaluators have been performing complexity-informed evaluations, like DE, for a long while, but "in the shadows (or even in the closet)" and that this is now changing with increasing definition and acceptance of such approaches (p. 193).

Mobilizing Evaluation Knowledge. Diverse participation, co-creative processes, and other features intrinsic to SI and SE are placing new demands on evaluators. Evaluation has an important role to play, but we need to be similarly innovative, flexible and creative in order to help innovators meet the challenges they face. We see the rich base in the evaluation literature as a valuable input to innovation in this space. Indeed, Vo et al. (2016) describe impact measurement (under which they include social accounting) and evaluation as "cognate practices" with "many parallel concepts" (p. 474). A lack of awareness or perceived lack of relevance of evaluation as a field to decision-making in SE contexts is concerning.

In like fashion, the SE literature points to a perception that the social sector is unresponsive, ineffective and inefficient and that there is not enough rigor in measurement to ensure accountability and drive change (Nicholls, 2009). Whether or not this is the case, weakening public trust in the social sector is an issue for evaluators. Yu & McLaughlin (2015) argue that decades of funding to NGOs has produced little evidence of impact, described as a "sobering reality" and leading to accountability of NGOs becoming a prominent debate among academics, politicians, the media and the public (pp. 24-26). SEs have been on a growth trend for at least a decade, and position themselves as a more effective alternative. Vo et al. (2016) connect the rise of social enterprise also to reduction in public funding of social services and social programs since the 2008 economic crisis, and continuing demand for privately-funded alternatives. It is likely that social impact measurement will continue to develop and increase in prominence alongside growth in SE.

SI and SE evaluation have been on parallel tracks, but may not be forever. It is important that evaluation professionals in both domains inform themselves to have this conversation and

to build on the strengths, opportunities and learning of both. The Impact Convergence

conference organized in tandem with AEA in 2016 is an excellent start. We also advocate more

attention be paid to mobilizing the evaluation knowledge base, created over 40 years of effort,

to other actors. This includes reaching out proactively to funding and commissioning

communities to engage them more closely in the conversation about lessons learned through

evaluation's long experience. Individual evaluators already do this, but their reach is limited.

We suggest this is a role for professional evaluation organizations to adopt as a voice of the

profession, and for their members to support. For example, stronger linkages could be made

with other groups of professionals such as the Social Impact Analysts Association (SIAA) and the

Social Investment Network (see Vo et al., 2016).

Implications for Research on Evaluation (RoE)

We also advocate a role for researchers on evaluation. There is a need for more RoE on

SI and SE evaluation, covering a greater diversity of approaches. Despite extensive search and

with key terms broadly defined, we found only 27 empirical studies on SI evaluation. We found

no studies on SE evaluation in program evaluation journals, although at least two (Cordes,

2016; Vo et al., 2016) have been published since. Also, while single case studies and case

narratives offer richness of detail, other types of designs and methodological perspectives are

needed. To support RoE in this domain, work to advance the clarity around concepts and

terminology of SI is needed. This is also needed for newer approaches in the SE domain, in the

vein of what has recently emerged for DE (Patton, McKegg & Wehipeihana, 2015). Research on

evaluation is essential to help evaluators consolidate, reflect on and advance practice in this

new space. Our findings invite reflection and debate about how evaluation is being undertaken

for social innovations and social enterprises, what knowledge base informs it, and how we

might support a similar spirit of innovation and collaboration in evaluation.
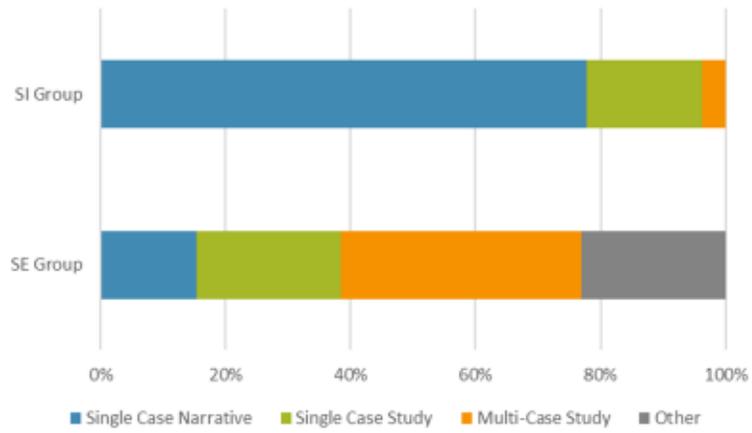
Tables and Figures

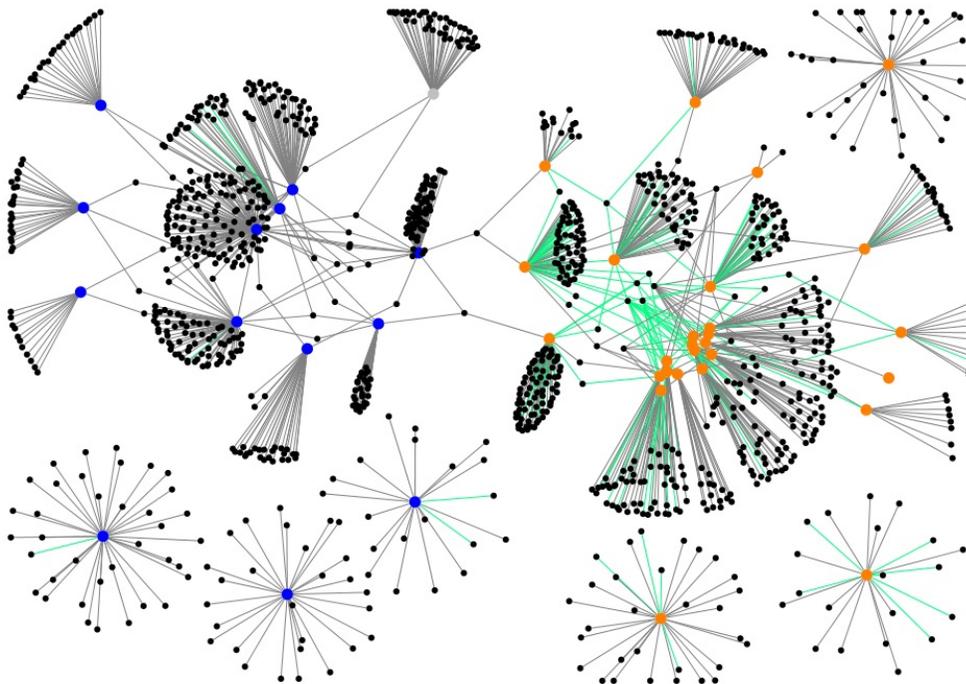Figure 1. Study designs of SI and SE groups in the sample.

Figure 2. Map of referenced source materials in the SI and SE groups of studies in the sample. Blue circles are articles in the SE group, orange circles are articles in the SI group. One grey circle is an article describing SE and non-SE evaluation. Black circles are references. Green lines are evaluation sources.
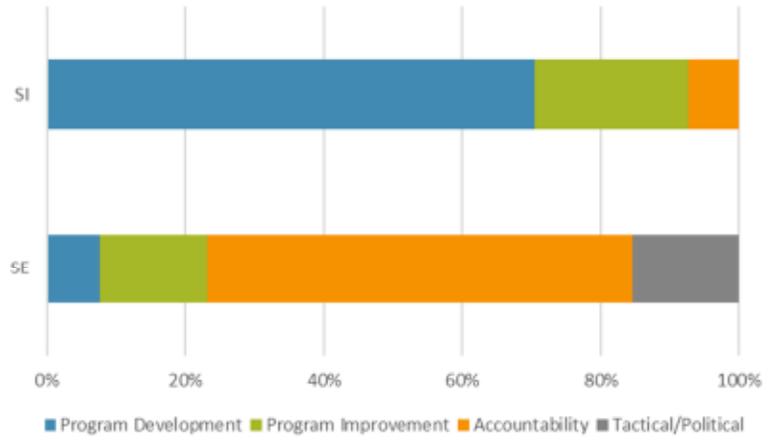
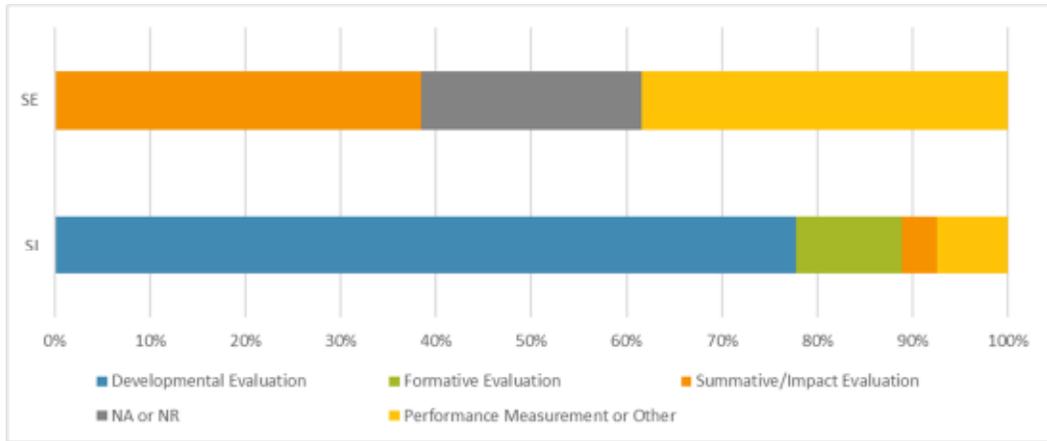**Figure 3. Primary reported purposes for evaluation.**
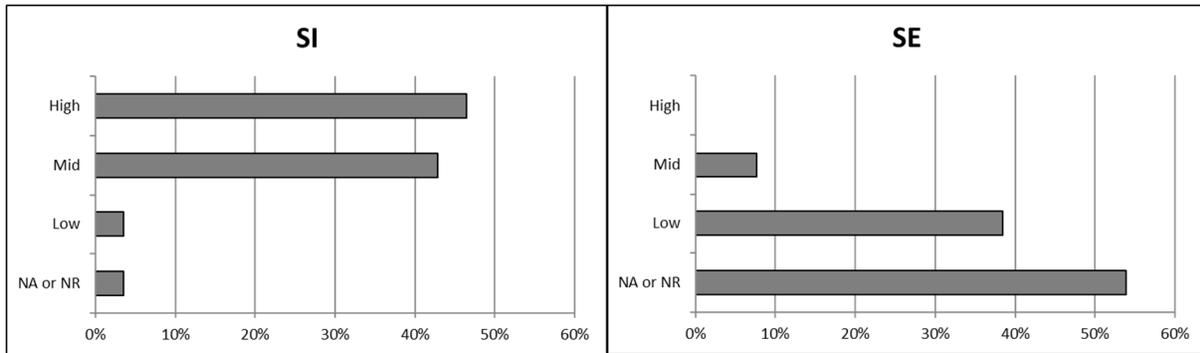

Figure 4. Evaluation or measurement 'types' reported

Figure 5. Percent of studies reporting different levels of collaborative practices.

Appendix A.

Table 1. SE studies reviewed[10]

Article Citation	Study Type	SE Focus; Geographic Location
Arasti, Z., Zarei, H., & Didehvar, F. (2015).	Qualitative, Interview	SE Policy; Iran
Arena, M., Azzone, G., & Bengo, I. (2015).	Case Study	Energy; Italy
Barraket, J & Yousefpour, N. (2013).	Multi-Case Study	Various; Australia
Galvin, M., & Iannotti, D. (2015).	Case Study	Agriculture; Tanzania, Kenya
Hazenberg, R., Seddon, F. & Denny, S. (2014).	Multi-Case Study	Work Integration/Training; UK
Jackson, E. & Tarsilla, M. (2012).	Case Study	Micro-Loan; Canada
Luke, B., Barraket, J. & Eversole, R. (2013).	Multi-Case Study	Various; Australia
Ndemo, E.B. (2006).	Qualitative, interview, ethnographic study	Various; Kenya
Nicholls, A. (2009).	Multi-Case Study	Various; UK
Palmas, K. & Lindberg, J. (2013).	Multi-Case Study	Agriculture; Sri Lanka
Pollard, N. & Cook, S. (2012).	Case Narrative	Mental Health; UK
Popielarski, J. & Cotugna, N. (2010)	Case Narrative	Poverty Reduction; USA
Sadownik, B. (2012).	Mixed Methods	Various; Canada
Yu, S., & McLaughlin, D. A. (2013)[11].	Multi-Case Study	International development; Canada

---

[10] For descriptions of the SI studies reviewed, see Milley et al., (2016).
[11] This study discusses both SE and non-SE evaluation. This article was included in qualitative discussion where relevant, but not included in the categorical data as it belongs to both groups.

References

Allen, S., Hunsicer, D., Kjaer, M., Krimmel, R., Plotkin, G. & Skeith, K. (2015). Adapted developmental evaluation with USAID's People-to-People Reconciliation Fund Program. In M. Patton, K., McKegg, & N. Wehipeihana (Eds.), *Developmental Evaluation Exemplars: Principles in Practice* (pp. 216-233). New York: The Guilford Press.

Antadze, N. & Westley, F. (2012). Impact metrics for social innovation: Barriers or bridges to radical change? *Journal of Social Entrepreneurship*, 3(2), 133-150.

Antadze, N. & Westley, F. (2010). Funding social innovation: how do we know what to grow? *The Philanthropist*, 23(3), 343-356.

Anzois, V. & Sardo, S. (2016). Dynamic Systems and the role of evaluation: the case of the green communities project. *Evaluation and Program Planning*, 54, 162-172.

Arasti, Z., Zarei, H., & Didehvar, F. (2015). Identifying the evaluative indicators of regulatory policies for the development of social entrepreneurship. *Public Organization Review*, 15(3), 453-474.

Arena, M., Azzone, G., & Bengo, I. (2015). Performance Measurement for Social Enterprises. *Voluntas: International Journal Of Voluntary & Nonprofit Organizations*, 26(2), 649-672.

Asher, J., Foote, N., Radner, J., & Warren, T. (2015). Science and how we care for needy young children: The Frontiers of Innovation Initiative. In M. Patton, K., McKegg, & N. Wehipeihana (Eds.), *Developmental Evaluation Exemplars: Principles in Practice* (pp. 103-124). New York: The Guilford Press.

Barraket, J & Yousefpour, N. (2013). Evaluation and Social Impact Measurement Amongst Small to Medium Social Enterprises: Process, Purpose and Value. *Australian Journal of Public Administration*, 72(4): 447-458.

Cabaj, M., Leviten-Reid, E., Vocisano, D. & Rawlins, M. (2015). An example of patch evaluation: Vibrant Communities Canada. In M. Patton, K., McKegg, & N. Wehipeihana (Eds.), *Developmental Evaluation Exemplars: Principles in Practice* (pp. 163-191). New York: The Guilford Press.

Cherniss, C. & Fishman, F. (2004). The Mesquite 'MicroSociety' school identifying organizational factors that facilitate successful adoption of an innovative program. *Evaluation and Program Planning*, 27(1), 79-88.

Chouinard, J. (2016). Introduction: decolonizing international development evaluation. *Canadian Journal of Program Evaluation*, 30(3), 237-247.

Chouinard, J. and Hopson, R. (2016). A critical exploration of culture in international development evaluation. *Canadian Journal of Program Evaluation*, 30(3), 248-276.

Cordes, J. (2016). Using cost-benefit analysis and social return on investment to evaluate the impact of social enterprise: promises, implementation, and limitations. *Evaluation and Program Planning*, in press.

Corporation for National and Community Service. (2016). About the Social Innovation Fund. Retrieved from http://www.nationalservice.gov/programs/social-innovation-fund/about-sif

Corporation for National and Community Service. (n.d.) Social innovation fund evaluation plan guidance: A step-by-step guide to designing a rigorous evaluation. Retrieved from https://www.nationalservice.gov/sites/default/files/documents/SIF%20Evaluation%20guidance%208%205%202014.pdf

Cousins, J.B. (2004). Commentary: Minimizing evaluation misuse as principled practice. *American Journal of Evaluation*, 25(3), 391-397.

Cousins, J.B., Goh, S.C., Elliott, C.J. and Bourgeois, I. (2014). Framing the capacity to do and use evaluation. In: J.B. Cousins and I. Bourgeois (Eds), Organizational capacity to do and use evaluation. *New Directions for Evaluation*, 141, 7–23.

Cousins, J.B. and Leithwood, K.A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, 56(3), 331-364.

Dart, R. (2004). The legitimacy of social enterprise. *Nonprofit Management and Leadership*, 14(4), 411–424.

Dickson, R., & Saunders, M. (2014). Developmental evaluation Lessons for evaluative practice from the SEARCH Program. *Evaluation*, 20(2), 176-194.

Dufour, S., Lessard, D., & Chamberland, C. (2014). Facilitators and barriers to implementation of the AIDES initiative, a social innovation for participative assessment of children in need and for coordination of services. *Evaluation and Program Planning*, 47, 64-70.

Ebrahim, A. & Rangan, V.K. (2014). What impact? A framework for measuring the scale and scope of social performance. *California Management Review*, 56(3), 118-141.

European Commission (2016) Social innovation. Available at: https://ec.europa.eu/growth/industry/innovation/policy/social_en (accessed 5 July 2016)

Galvin, M., & Iannotti, D. (2015). Social Enterprise and Development: The KickStart Model. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 26(2), 421-441.

Gamble, J., Van Sluys, S. & Watson, L. (2015). Fostering learning through developmental evaluation with a nontraditional arts organization and a traditional community funder. In M. Patton, K., McKegg, & N. Wehipeihana (Eds.), *Developmental Evaluation Exemplars: Principles in Practice* (pp. 83-102). New York: The Guilford Press.

Gopal, S., Mack, K., & Kutzli, C. (2015). Using developmental evaluation to support college access and success Challenge Scholars. In M. Patton, K., McKegg, & N. Wehipeihana

(Eds.), *Developmental Evaluation Exemplars: Principles in Practice* (pp. 45-62). New York: The Guilford Press.

Government of Canada. (2014). Budget 2014. Retrieved from http://www.budget.gc.ca/2014/docs/plan/ch3-2-eng.html

Hazenberg, R., Seddon, F. & Denny, S. (2014). Investigating the outcome performance of work-integration social enterprises (WISE's). *Public Management Review*, 16(6), 876-899.

Howard, J., Flores, A., & Hambleton, R. (2015). Evaluation as the co-construction of knowledge Case studies of place-based leadership and public service innovation. *Advances in Program Evaluation*, 15, 207-227.

Jackson, E. & Tarsilla, M. (2012). Mixed methods in social accounting: Evaluating the micro-loan program of Alterna Savings Credit Union. In L. Mook (Ed.) *Accounting for Social Value* (pp. 139-166). Toronto: University of Toronto Press.

King, J.A. (1988). Research on evaluation use and its implications for evaluation research and practice. *Studies Educational Evaluation*, 14(3), 285-299.

Kuji-Shikatani, K., Gallagher, M.J., Franz, R. & Borner, M. (2015). Leadership's role in building the education sector's capacity to use evaluative thinking: The example of the Ontario Ministry of Education. In M. Patton, K., McKegg, & N. Wehipeihana (Eds.), *Developmental Evaluation Exemplars: Principles in Practice* (pp. 252-270). New York: The Guilford Press.

Lam, C. & Shulha, L. (2015). Insights on Using Developmental Evaluation for Innovating A Case Study on the Cocreation of an Innovative Program. *American Journal of Evaluation*, 36(3), p. 147-169.

Langlois, Blanchet-Cohen, Beer (2013). The art of the nudge Five practices for developmental evaluators. *Canadian Journal of Program Evaluation*, 27(2): 39-59.

Luke, B., Barraket, J. & Eversole, R. (2013). Measurement as legitimacy versus legitimacy of measures. *Qualitative Research in Accounting and Management*, 10(3/4): 234-258.

Marcellus (2004). Developmental evaluation of the Safe Babies projects: Application of the COECA model. *Issues in Comprehensive Pediatric Nursing*, 27, 107–119.

Mathie, A., & Peters, B. (2014). Joint (ad)ventures and (in)credible journeys evaluating innovation asset-based community development in Ethiopia. *Development in Practice*, 24(3), 405-419

McKegg, K., Wehipeihana, N., Becroft, M. & Gill, J. (2015). Developmental evaluation's role in supporting community-led solutions for Mauri and pacific young people's educational success The Foundation North Maori and Pacific Education Initiative. In M. Patton, K., McKegg, & N. Wehipeihana (Eds.), *Developmental Evaluation Exemplars: Principles in Practice* (pp. 125-142). New York: The Guilford Press.

Milley, P., Szijarto, B., Svensson, K., & Cousins, J.B. (2016, September). *The evaluation and social innovation connection: A systematic review of empirical evaluation studies conducted in*

social innovation contexts, 2000-2015. Paper presented at the European Evaluation Society Conference, Maastricht, NL.

Moore, M. & Cady, J. (2015). Developmental evaluation in the McKnight Foundation's Collaborative Crop Research Program A journey of discovery. In M. Patton, K., McKegg, & N. Wehipeihana (Eds.), *Developmental Evaluation Exemplars: Principles in Practice* (pp. 143-162). New York: The Guilford Press.

Murphy, N. (2015). Nine guiding principles to help youth overcome homelessness A principles-focused developmental evaluation. In M. Patton, K., McKegg, & N. Wehipeihana (Eds.), *Developmental Evaluation Exemplars: Principles in Practice* (pp. 63-82). New York: The Guilford Press.

Ndemo, E.B. (2006). Assessing sustainability of faith-based enterprises in Kenya. International *Journal of Social Economics*, 33(5-6): 446-462.

Nicholls, A. (2008). *Social entrepreneurship: New models of sustainable social change*. Oxford: Oxford University Press.

Nicholls, A. (2009). We do good things, don't we?: 'Blended value accounting' in social entrepreneur- ship. *Accounting, Organizations and Society*, 34(6–7), 755–769.

Palmas, K. & Lindberg, J. (2013). Livelihoods or ecopreneurship? Agro-economic experiments in Hambantota, Sri Lanka. *Journal of Enterprising Communities*, 7(2), 125-135.

Patton, M. (2011). Developmental evaluation: Applying complexity concepts to enhance innovation and use. New York: Guilford Press.

Patton, M., McKegg, K. & Wehipeihana, N. (2015). *Developmental evaluation exemplars: principles in practice*. New York: The Guilford Press.

Pollard, N. & Cook, S. (2012). The power of low-key groupwork activities in mental health support work. *Groupwork*, 22(3), 7-32.

Popielarski, J. & Cotugna, N. (2010) Fighting hunger through innovation: Evaluation of a food bank's social enterprise venture. *Journal of Hunger & Environmental Nutrition*, 5(1), 56-69.

Poth, C., Pinto, D., Howery, K. (2012). Addressing the challenges encountered during a developmental evaluation Implications for evaluation practice. *Canadian Journal of Program Evaluation*, 26(1), 39- 48.

Preskill, H., & Beer T. (2012). *Evaluating social innovation. Washington: FSG and the Centre for Evaluation Innovation*. Retrieved from: http://www.evaluationinnovation.org/sites/default/files/EvaluatingSocialInnovation.pdf

Preskill, H., Holladay, R., Gopal, S., Kutzli, C., Mack, K., Jhawar, M., Cook, J., Radner, J., Devji-Esmail, A. (2014). *Evaluating complexity while creating equitable and sustainable systems change*. Paper presented at the American Evaluation Association conference, Denver, CO, USA.

Ramirez, R., Kora, G., & Shephard, D. (2015). Utilization Focused Developmental Evaluation Learning Through Practice. *Journal Of MultiDisciplinary Evaluation*, 11(24), 37-53.

Ramstad, E. (2009). Developmental evaluation framework for innovation and learning networks Integration of the structure, process and outcomes. *Journal of Workplace Learning*, 21(3) 181-197.

Rey, L., Tremblay, M.C., & Brousselle, A. (2014). Managing tensions between evaluation and research Illustrative cases of developmental evaluation in the context of research. *American Journal of Evaluation*, 35(1), 45-60.

Rogers, Y. (2008, August 11). New ways of doing. *The Guardian*. Retrieved from https://www.theguardian.com/commentisfree/2008/aug/11/welfare.health

Saari & Kallio (2011). Developmental impact evaluation for facilitating learning in innovation networks. *American Journal of Evaluation*, 32(2), 227-245.

Sadownik, B. (2012). The demonstrating value initiative: Social accounting for social enterprises. In L. Mook (Ed.) *Accounting for Social Value*, (pp. 139-166). Toronto: University of Toronto Press.

Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: theory, research, and practice since 1986. *Evaluation Practice*, 18(3), 195-208.

Social Science and Humanities Research Council. (2016). *Community and college social innovation fund*. Retrieved from http://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/social_innovation-innovation_sociale-eng.aspx

Svensson, K., Szijarto, B., Milley, P., & Cousins, J.B. (2016, October). *Evaluating social innovations: Implications for evaluation design*. Paper presented at the American Evaluation Society Conference, Atlanta, USA.

Tan, E., McGill, S., Tanner, E., Carlson, M., Rebok, G., Seeman, T. & Fried, L. (2014). The evolution of an academic-community partnership in the design, implementation, and evaluation of experience corps Baltimore city a courtship model. *The Gerontologist*, 54(2), 314-321.

TEPSIE. (2014). *Social Innovation Theory and Research: A Summary of the Findings from TEPSIE*. A deliverable of the project: 'The theoretical, empirical and policy foundations for building social innovation in Europe' (TEPSIE), European Commission – 7th Framework Programme. Brussels: European Commission, DG Research. Retrieved from: http://www.tepsie.eu/images/documents/research_report_final_web.pdf

Togni, S., Askew, D., Rogers, L., Potter, N., Egert, S., Hayman, N., Cass, A., & Brown, A. (2015). Creating safety to explore Strengthening innovation in an Australian indigenous primary health care setting through developmental evaluation. In M. Patton, K. McKegg, & N. Wehipeihana (Eds.), *Developmental Evaluation Exemplars: Principles in Practice* (pp. 234-251). New York The Guilford Press.

Vo, A., Christie, C. A. & Rohanna, K. (2016). Understanding evaluation practice within the context of social investment. *Evaluation*, 22(4), 470-488.

Wehipeihana, N., McKegg, K., Thompson, V. & Pipi, K. (2015). Cultural responsiveness through developmental evaluation Indigenous innovations in sport and traditional Maori recreation. In M. Patton, K. McKegg, & N. Wehipeihana (Eds.), *Developmental Evaluation Exemplars: Principles in Practice* (pp. 25-44). New York The Guilford Press.

Weiss, C.H. (1997). Have we learned anything new about the use of evaluation? *American Journal of Evaluation,* 19(1), 21-33.

Weiss, C.H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice*, 9(1), 5-19.

Westley, F. (2013). *The history of social innovation*. Keynote speech at NESTA Social Frontiers Conference, November 14, London. Retrieved from: https://www.nesta.org.uk/sites/default/files/key_note_speech_frances_westley_on_the_history_of_social_innovation.pdf

Wilson-Grau, R., Kosterink, P. & Scheers, G. (2015). Outcome Harvesting A developmental evaluation inquiry framework supporting the development of an international social change network. In M. Patton, K. McKegg, & N. Wehipeihana (Eds.), *Developmental Evaluation Exemplars: Principles in Practice* (pp. 192-215). New York The Guilford Press.

Yu, S., & McLaughlin, D. A. (2013). Program Evaluation and Impact Assessment in International Non-Governmental Organizations (INGOs) Exploring Roles, Benefits, and Challenges. *Canadian Journal of Nonprofit & Social Economy Research*, 4(2), 23-38.

Zamir, J. & Jaber A. (2015). Promoting Social Justice Through a New Teacher Training Program for the Bedouin Population in the Negev An Evaluation Case Study. *New Directions for Evaluation*, 146, 71-82.