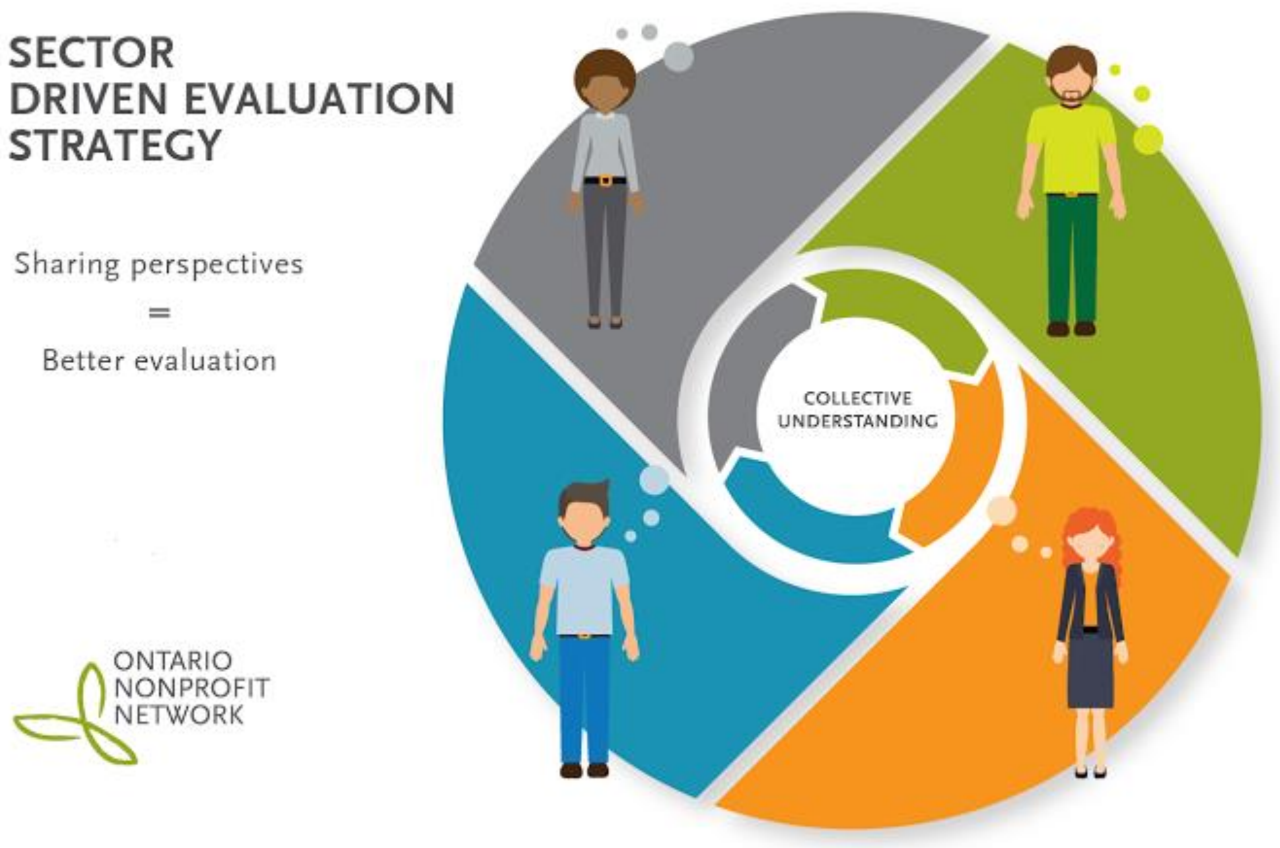
720 Bathurst Street, Suite 405
Toronto, Ontario
M5S 2R4

ONTARIO
NONPROFIT
NETWORK

(416) 642 - 5786
admin@theonn.ca
theonn.ca



Evaluation Literature Review

January 2016 | Ontario Nonprofit Network

Andrew Taylor

Ben Liadsky

About the Authors

Andrew Taylor is co-owner of Taylor Newberry Consulting, a Guelph-based firm that specializes in developing research and evaluation solutions for public sector organizations. He is also ONN's Resident Evaluation Expert. He has helped organizations across Canada develop impact strategies and measurement systems that are evidence based, manageable, and meaningful. He thinks evaluation is only useful if it answers questions that matter and enables people to act in new ways.

Ben Liadsky is ONN's Evaluation Program Associate. He has more than five years of experience working in the nonprofit sector in a variety of capacities from project management to fundraising to communications. He believes it's time to change the conversation around evaluation.

About ONN

Organized in 2007 and incorporated as a nonprofit in 2014, the Ontario Nonprofit Network (ONN) is the convening network for the approximately 55,000 nonprofit and charitable organizations across Ontario. As a 7,000-strong provincial network, with a volunteer base of 300 sector leaders, ONN brings the diverse voices of the sector to government, funders and the business sector to create and influence systemic change. ONN activates its volunteer base and the network to develop and analyze policy, and work on strategic issues through its working groups, engagement of nonprofits and charities and government.

Our Vision

A Strong and Resilient Nonprofit Sector. Thriving Communities. A Dynamic Province.

Our Mission

To engage, advocate, and lead with—and for—nonprofit and charitable organizations that work for the public benefit in Ontario.

Our Values

Courage to take risks and do things differently.
Diversity of perspectives, **creativity** and **expertise** to get stuff done.
Optimism and **determination**.
Solutions created by the sector, with the sector, for the sector.
Celebrating our successes and **learning** from our experiences.
Strength that comes from working **together**.

Ontario Nonprofit Network
720 Bathurst Street, Suite 405
Toronto, Ontario
M5S 2R4

Website: theonn.ca **Email:** mailto:info@theonn.ca **Tel:** (416) 642-5786

Table of Contents

Acknowledgements

While no literature review is ever perfect, the process of researching and writing this report has been an enlightening experience. When this review was first discussed, it was intended to be a relatively simple endeavour designed to showcase a few key sources and provide some commentary as a way to help us develop context and start to frame up our thinking. However, it quickly became apparent that this document could serve a broader purpose and could become a foundational piece for ONN's evaluation project. As such, it has helped us to refine our thoughts and focus in on some of the systemic issues of evaluation in the nonprofit sector. As it evolved, in true ONN fashion, we began reaching out to our network for feedback. While much of the content comes from digging through numerous sources — academic journals, websites, online blogs, and various funder and nonprofit reports — the work is also grounded in the conversations held with different people. For this reason, we would like to acknowledge the support and feedback from all those who contributed their time in talking with us and, in particular, we wish to acknowledge the members of the ONN Evaluation Advisory Committee for their valuable insights. Going forward, we look to build on this document by continuing to reach out to the sector in creating a Sector Driven Evaluation Strategy.

Introduction

The Ontario Nonprofit Network (ONN) is developing a Sector Driven Evaluation Strategy to empower nonprofits to become more actively involved in setting the evaluation agenda. We are focused on how evaluation is negotiated at a high level and making sense of evaluation systems. By looking at the complete picture of evaluation in the sector, with an eye and an ear to the needs of different stakeholders, our hope is to be able to design a strategy that helps nonprofits to make evaluation work for them.

In Ontario's nonprofit sector, evaluation is least likely to lead to positive action when it is used only as a means to hold nonprofits accountable for their use of grant money. The potential for learning and action is even lower if the process is poorly explained, based on unrealistic expectations, or under-resourced. This is not in the least surprising. The evaluation literature has identified these mistakes and their implications time and again. It has also developed a wide range of ideas about why these problems occur and how to avoid them. The Sector Driven Evaluation Strategy is designed to make evaluation less about paperwork and more about insight.

Finding ways to make evaluation more meaningful and more useful has been a key theme in the evaluation literature since the discipline began, and there is no shortage of discussion around improving evaluation among nonprofit practitioners. The topic has been a highlight at ONN's annual conference in recent years. However, much of the discussion around improving evaluation focuses on methodology, tools, and indicators. There has been less attention paid to who is asking and determining the questions of evaluation, such as *who evaluation is for* and *what is its purpose*. Consequently, the purpose of this background paper is to review the literature on evaluation use with a particular focus on systemic factors. In other words, we are interested in looking at the relationship between evaluation practice and the overall structure and function of the nonprofit sector in Ontario.

We're interested in the policies and regulations that guide us, the roles played by various actors, the assumptions we make, the language we use, and the ways in which resources move through the sector. We're examining the purposes that evaluation serves, both overt and implicit. We want to learn more about the factors that make evaluations really useful, the issues that can get in the way of evaluations being useful, and ideas for improvement. Ultimately, our goal in this paper is to generate a broad vision to inform our project's final outcomes.

Evaluation is a term that is used in a variety of ways. In this paper, we use the term inclusively and we intend to cover performance measurement, evaluation of individual programs, and systems-level evaluation. It should also be noted that the body of literature is quite diverse and that evaluation as an issue does not exist in isolation from other areas of nonprofit work, such as fundraising/grant-writing and data collection and analysis. While these intersections are important, they are worthy of their own study and only briefly addressed in this document.

This paper is divided into five parts. Part one looks at the different types and purposes of evaluation and other measurement and accountability tools. Part two looks at the factors that contribute to making evaluations useful. Part three surveys what is happening in Ontario's nonprofit sector and identify some of the tensions and challenges that get in the way of useful evaluation. Part four offers up ideas on promoting increased use of evaluation. Finally, part five covers some early ideas from our research to date for potential strategies and solutions to be included in a Sector Driven Evaluation Strategy.

What is in scope?

The purpose of this project is to develop a sector-driven evaluation agenda. The literature on evaluation practice, evaluation use, and the factors that promote use is vast, so this first section has tried to identify our scope. The term *evaluation* can refer to a number of different kinds of measurement work. Of particular interest in this paper are performance measurement, program evaluation, systems evaluation and applied research. In this section the differences between these four types of work are explained. All four types are considered in the sections that follow and we have tried to be clear about which ones we are talking about at any given point.

1. Defining Evaluation

In the nonprofit sector, the term *evaluation* is used to cover a wide range of social research activities, undertaken by different stakeholder groups. For some, evaluation might mean a group of staff getting together at the end of a program cycle to reflect on how it went. For others, evaluation could be a complex, multi-year research project with sites all over the province and access to a large team of academic experts. Evaluation can also be undertaken for many different reasons. Sometimes, it is motivated by a desire to hold nonprofits accountable for their use of public money. At other times, ongoing program improvement, buy-in from partners, or program expansion might be the goal. Often, a single evaluation project has multiple goals.

Many in the field have pointed out that there really is no single definition that adequately captures all of this diversity.[i] When pressed, most authors in the field respond with something like Patton's definition: "Evaluation is the systematic collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding."[ii]

Evaluation is a relatively young discipline and has changed a lot over the years. Today, more nonprofits than ever report engaging in some form of evaluation,[iii] and the practice of evaluation has become more diverse.

Evaluation can help a nonprofit make sense of what it does and how it does it. It provides an opportunity to engage with all stakeholders, reflect on both failures and successes and learn from them in order to make evidence-based decisions. **Ultimately, one of the defining qualities of evaluation — and the quality that distinguishes it most clearly from applied research — is that it is designed to inform action in a direct, specific, and timely way.**

There are numerous ways in which evaluation findings can be used.[iv] The following table (Table 1) outlines a few.

Table 1. Evaluation uses.

Instrumental use	When evaluation recommendations are acted upon directly.
Conceptual use	When an evaluation helps users develop a deeper understanding of key issues or ideas. An evaluation report may, for example, help an organization to realize that board members and front line staff do not understand the principle of inclusion in the same way.
Symbolic, legitimative or *persuasive use*	When evaluation findings help to bolster the case for an intervention that is already planned, or to adjust the strategy for implementation of that intervention. An organization may already have a strong belief, based on years of experience, that a new strategy will work. An evaluation of a small pilot project may then help it get the resources to implement that strategy on a larger scale.
Process use	When the evaluation experience itself leads to action. A board member may sit in on a focus group, for example, and develop a better relationship with service users as a result (even before an evaluation report is released). When those involved in an evaluation project have a narrow conception of use, they miss important insights.

Indeed, there are many different approaches to evaluation and the field is evolving all the time.[v] For the purposes of this paper, there are a few definitional issues that are particularly salient. They are explored in the subsections below.

Evaluation and performance measurement

The simplest form of evaluation is often called *performance measurement* or program monitoring. This is the ongoing, day-to-day data gathering that program staff and volunteers do as part of their job. It tends to use low-cost, less intensive data gathering techniques. It often focuses on tracking program processes and outputs (e.g., attendance rates, demographic information about participants or basic feedback on satisfaction). Performance measurement often generates data that managers can use quickly and frequently, and as a result it helps in monitoring the program. Accreditation processes often rely heavily on performance measurement data, as do quality improvement processes.

Performance measurement work is so integrated into the process of program management, its purpose is often explicitly tied to accountability.[vi] Small organizations often have home-grown performance measurement systems that they have created internally. In large organizations, performance measurement tools are often integrated into professionally designed database systems so that information can be gathered, reviewed and interpreted quickly. Sometimes, entire systems or networks share a single performance measurement database. Ontario's Community Health Centres share a performance measurement system, for example.

Program evaluation work tends to be more intensive, more formal, and more time-limited than ongoing program monitoring efforts. While program monitoring typically produces a simple summary of key statistics or a dashboard, a program evaluation project typically begins with a critical analysis of the theoretical assumptions underlying a program (using a theory of change) and produces an analytical report with conclusions and recommendations.[vii] Program evaluation work often involves deeper investigation into the outcomes or impacts as well as questions about process. It attempts to determine whether a program led to change and why.

It can be challenging to draw a clear distinction between the two forms of measurement. Monitoring and accountability are often the primary purposes of performance measurement work. While program evaluation work is also used for accountability purposes, it typically has other goals as well. It is intended to lead to on-the-ground insights, improvements in programming and even increase buy-in from various stakeholders. [viii]

Evaluation and applied research

While evaluation and research are both ways of better understanding social issues, *applied research* is usually intended to generate new knowledge for a wide audience. The Better Beginnings, Better Futures longitudinal research project is an excellent example of an important applied research project in the nonprofit sector.[ix]

As compared to evaluation, applied research is often more time consuming, more theory-driven, and more expensive. It often involves consideration of data from multiple programs or program sites. It is typically designed and carried out by academic researchers who are content experts in the research topic under consideration. Its primary purpose is to create a new, generalized knowledgebase and therefore, it may not always generate practical recommendations for immediate local action.

Program evaluation, however, is usually designed to generate concrete, practical recommendations for people running a specific program or system. Evaluation is distinct from applied research in that it considers local context, the values of the people involved, program side effects, and other factors that an applied researcher would not consider, in order to generate "evaluative conclusions."[x]

Program evaluation and systems-level evaluation

While evaluation techniques can be used on a small scale to study an individual program or intervention within one community, one can also evaluate entire organizations, or even groups of organizations or large systems.[xi] For example, a government may simultaneously invest in life skills training programs for youth in detention centres and community integration supports for these same youth after they leave custody. The evaluation of these two interventions is necessarily intertwined. Complex service interventions are very likely to work differently in different communities and to evolve over time.[xii] One city may choose to integrate community integration supports for youth leaving detention with programs for other groups of at risk youth, while another may not. Best practices developed in one community may eventually be adopted on a larger scale.

Perhaps the most important defining feature of complex interventions is that they require and encourage many different people, at different points along the "chain of intervention," to be actively involved in various kinds of data gathering (including performance measurement, program evaluation and applied research). These diverse groups of people will inevitably have different priorities for the questions that they want answered through evaluation. Accountability in these situations is often dense and non-linear, with different funders and different levels of government asking related but distinct questions at different times, and requiring different types of data.[xiii]

When governments or other large funders commission a large number of related evaluation projects, then seek to combine the findings of these projects into a larger report on the impact of a complex and diverse set of community investments, they are undertaking a process that is far more complex than ordinary evaluation. They are also asking very different kinds of questions than local agencies might ask. They are often focused on keeping the system running smoothly and managing accountability. At best, they are also interested in how the various parts of these interventions work together. Rarely are they interested in the immediate contribution of their programs to the lives of a specific, localized group of users.

What's the right measurement approach for the job?

Any organization that uses public resources faces pressure to demonstrate that it has used those resources responsibly. This is why program evaluation and related techniques are such important tools for Ontario's nonprofit sector. Understanding the distinctions between different approaches can help to manage expectations. If a government department wants to identify best practices for promoting physical activity for seniors or generate rigorous evidence that safe injection sites reduce deaths, an applied research project may be more appropriate than a program evaluation. If a small nonprofit can only afford a very simple performance measurement system that tracks attendance and demographics, it may not be reasonable to expect it to report on client outcomes. Tables 4 and 5 in Appendix C provide a summary of these distinctions.

In essence, knowing which measurement approach is right for the job is critically important for both nonprofits and funders to make the most of their time and money. How to choose which approach is right though is not always an easy decision. As the next section illustrates, in order to make evaluation useful, there are number of factors to consider.

2. Making evaluation useful

In part one, it was stated that a focus on use is a key distinguishing feature of evaluation work. *Utilization-Focused Evaluation* is an approach, designed by Michael Quinn Patton, which tries to maximize the potential for action by identifying the intended uses of an evaluation in the planning stages.[xiv] Patton draws a distinction between the *audience* for a report (all those who might have a passive interest in reading a report) and the *intended users* (those who are expected to take specific types of action on the basis of the evaluation). Uses and users are to be named in the most specific terms possible (e.g., "the executive director will use the findings from this evaluation to prepare her report on progress against the strategic plan" is better than "the evaluation will be used by organizational leadership for planning"). In simple terms, no evaluation project should begin unless the individuals who will act upon it have been clearly identified and unless those people have made a personal commitment to using the findings.[xv]

Evaluators often make the mistake of rushing to considerations of methodology before working to make sure that the people involved are ready for evaluation and capable of benefitting from it.[xvi] Key steps in promoting use include: overcoming apprehension about evaluation; generating evaluation questions that are meaningful to all stakeholder groups; and convincing those involved that there is a strong commitment to use from the beginning of the process.[xvii]

In a review focused specifically on utilization in the context of the funder-fundee relationship, a number of factors were identified (and outlined in Table 2) that have an impact on the likelihood of evaluation use.[xviii]

Table 2. Factors that predict whether evaluation will be used.

Contextual factors	*Human & relationship factors*	*Evaluation factors*
● money and resources ● program stability ● skill and experience of program staff	● skill and experience of evaluators ● commitment to translate the evaluation into action ● engagement and commitment among intended users of the evaluation ● trust and rapport between stakeholder groups involved	● user involvement in evaluation design ● relevance and usefulness of evaluation questions ● communication that is timely, transparent, honest, credible, and inclusive of all important stakeholders

In the end, human and relationship factors prove to be the single most important forces in utilization (much more important than the quality of the evaluation methodology or the depth of the analysis, for example) and they continue to play this role throughout all phases of the evaluation process from planning, through data collection and into analysis and reporting.[xix] When a strong and trusting relationship exists, evaluations become easier to manage, less costly, and more focused, as well as more useful.[xx]

Complex interventions and useful evaluation

Facilitating good communication across stakeholders becomes more difficult and more important as evaluation work becomes more complex.[xxi] In very complex systems, this type of facilitation can become the largest component of an evaluator’s role. Conflict management and consensus building skills become essential. Evaluation efforts are often initiated by different *users* for different, sometimes conflicting reasons. While difficult, taking the time to find the common interest across many different stakeholder groups is key and investing in ongoing communication at all levels throughout the evaluation process becomes even more important.[xxii]

A few years ago, the Treasury Board of Canada studied factors that lead to effective evaluation in a federal government context. Support from senior management and a participatory approach to evaluation planning were found to be key predictors of how useful an evaluation project would be. They also identified a number of more specific best practices for evaluation of complex government initiatives, including:

- involvement of the front line program managers (as well as high-level senior managers) in all aspects of evaluation design, including terms of reference, choice of questions, and tool design;
- mutual agreement on the evaluation objectives and the criteria against which evaluation success will be judged, achieved through in-depth discussion about expectations among program staff, government funders, managers, directors and evaluators;
- frequent, open, rapid communication among all stakeholders including clients. Repeated check-ins with front-line program managers and other intended users throughout the process to ensure that the evaluation was still meeting expectations and addressing concerns; and
- collaboration around the development and presentation of the organization's response to evaluation findings.[xxiii]

In short, the Treasury Board has learned that front-line partners are almost always high on the list of intended evaluation users, even when the evaluation is focused on high-level systems questions. Executive directors and managers within nonprofits are often the ones with the most influence on how an evaluation will be implemented and used.[xxiv] Research suggests that nonprofit leaders are willing to try to embrace the findings of an evaluation and to learn from them.[xxv] It is for this reason that progressive grantmakers are increasingly interested in promoting a *culture of learning* whereby nonprofits use evaluation as a way to reflect, learn, and share their failures and successes.[xxvi] In this sense, evaluation is something that is embraced by everyone within the nonprofit as a principle and, therefore, a commitment to ongoing evaluation becomes part of organizational culture within the nonprofit.

Reflections on evaluations that get used

Evaluation exists in order to inform action. The evaluation strategies that accomplish this most effectively are not necessarily those that have complex data or sophisticated methodologies. The most effective strategies invest time in building shared ownership and commitment to act. The resources reviewed in this section make it clear that choosing the right tool is a decision that is best made collaboratively**.** While a simple performance measurement system may require less intensive engagement with stakeholders than a more complex, multi-method evaluation project, its success still depends on a shared understanding of the purpose of performance measurement. Buy-in from front-line staff may be even more important in the context of performance measurement than it is for more sophisticated forms of evaluation, since they hold primary responsibility for data gathering.

A way to keep use in focus

Given that the intention for use is one of the distinguishing features of evaluation compared to reporting or research, the Utilization-Focused Evaluation (UFE) approach is very helpful in framing up the questions under discussion in this paper. In particular, UFE helps to highlight the fact that the focus here is not on the technical aspects of measurement, but on the human and relationship factors that seem to be the best predictors of evaluation use.

3. Evaluation in the nonprofit sector: diagnosing the problem

A great deal of performance measurement and evaluation work takes place in Ontario's nonprofit sector, and it is clear that this kind of work has the potential to do a great deal of good. There are many organizations and networks that now act in a way that is more evidence based and grounded in the experiences of their clients as a result of their evaluation work. Some of these success stories are small and local, while others have led to significant change at a province-wide level. A recent American study found that 95% of nonprofits engage in some form of evaluation, and most use that information to improve existing programs, report to the board, and plan for the future.[xxvii] There is some evidence that interest in evaluation work is increasing among nonprofits and that evaluation work is being used more often for a wider range of purposes.[xxviii]

That said, many feel that evaluation does not lead to action as often as it should. In 2014, one study found that 49% of American grantmakers shared their evaluation findings with other grantmakers, while 46% reported back to grantees, and 20% used evaluation findings in efforts to influence public policy or government funding.[xxix] However, many still find the process frustrating, stressful, and, at times, fruitless or even damaging to their communities. The purpose of this section is to try to better understand why this is the case.

Diagnosing the problem of underutilization

The research literature identifies a number of factors that may explain why evaluation in Ontario's nonprofit sector often becomes frustrating and ends up being under-used.

- A focus on accountability over learning
- A mismatch between approach and expectations
- A mismatch between investment and expectations
- Inadequate communication
- Inattention to certain kinds of evaluation questions

This section reviews each of these factors in turn.

A focus on accountability can become an obstacle to utilization.

When nonprofits express frustration with evaluation, they tend to focus on situations in which evaluation work is required by an external funder. More specifically, challenges arise when funders are using evaluation methods as a means of holding the nonprofit accountable.

Some studies conclude that nonprofits are spending more and more resources on funder-mandated evaluation.[xxx] There is a strong perception in Ontario that this particular type of evaluation is becoming more common.[xxxi] This accountability shift reflects a larger societal change happening in countries around the world dating back to the late 1970s and the emergence of the New Public Management (NPM) framework.[xxxii] One of the features of NPM was a change in how accountability was defined. This change has seen "...accountancy expanded from checking whether the money had been spent in accordance with the rules, to checking on efficiency and effectiveness."[xxxiii]

Others have noted that accountability measures today "...tend to drive an ethics of accountability that is responsive largely to technical criteria and state-defined targets rather than an ethic of care concerned with very specific service user needs."[xxxiv] As a result, nonprofits feel ill-equipped to make strategic, informed decisions about evaluation.

In the early 2000s, an effort was made through the Voluntary Sector Initiative to strengthen relationships between nonprofits and the federal government, which included work on accountability, but progress was limited.[xxxv] As well, in the early 2000s, some prominent government spending scandals meant an increase in evaluation requirements for all those receiving government funds, including nonprofits, in order to demonstrate accountability.[xxxvi] More broadly, the emergence of charity watchdogs and media headlines on further scandals has put even more onus and attention on nonprofits and the work that they do.[xxxvii] Accordingly, evaluation and accountability have grown increasingly linked in the minds of many.

It is important to say that funder-driven, accountability-focused, performance measurement style evaluation is not the only kind of evaluation work taking place in Ontario's nonprofit sector. In fact, some research suggests that this type of evaluation is already on the wane. In 2003, Imagine Canada published a report suggesting that most evaluations in the sector (73%) were initiated by nonprofits themselves for internal reasons, while only 11% were initiated at the behest of funders.[xxxviii] In 2012, an American survey found that the organization's senior managers or CEO were a "primary audience" for 74% of evaluation projects, and that the board were a primary audience 65% of the time. Funders were a primary audience in 54% of cases.[xxxix] These figures vary by size of organization and by sub-sector, but suggest that there is much evaluation practice in the sector that falls outside of the problem that is being discussed in this paper.

While there may be debates about how common funder-driven, accountability-focused evaluation has become, there is consensus that it creates serious challenges for utilization. One of the effects of an evaluation system driven by funder accountability is that the process of evaluation itself comes to be seen as stressful, highly complex, and risky.[xl] "The net effect is that nonprofits are incented to focus on accommodating funders' growing demands for metrics that underscore their grantees' tactical prowess, efficiency and fiscal solvency, at the expense of addressing and solving complex social problems and concerns."[xli] When nonprofits are not actively involved in setting the evaluation agenda, they may have anxiety about intended uses. This leads to a disconnect between the funder and grantee, whereby the funder and nonprofit have different views and understanding of what should be measured and how.[xlii] As a result, nonprofits often feel as though they were "reeling from an accountability regime gone mad."[xliii]

In practical terms, this may lead to nonprofits tailoring their evaluation results to show a positive impact fearing (*Look-Good-Avoid-Blame* mindset) that a negative finding will lead to a funding cut.[xliv] In some cases, an evaluation report may be written but quickly becomes forgotten because it is unclear how it would help a nonprofit in its work.[xlv] Moreover, there may not be time or financial resources dedicated to the thinking and planning that would be required to make the shift to active use. Alternatively, a report may be produced, but it is not shared or used by all relevant stakeholders.[xlvi]

There is a mismatch between approach and expectations.

Funder-driven accountability evaluation tends to use an approach that aims to answer questions beyond its abilities.

Evaluations initiated by funders for the purposes of grant accountability often use the methods and approaches of performance measurement. These projects often use simple methods that focus on tracking outputs. These methods tend to be integrated in the process of program management, and be short and inflexible. Little or no resources are devoted to building a shared understanding of the assumptions and roadmap toward achieving long-term goals (theory of change) or reaching consensus on the intended use of results, because the intended uses are seen to be simple and straightforward.

The focus on performance measurement may arise from sincere intentions on the part of the funder. It may stem from a desire to reduce the burden placed on grant recipients. A measurement system that focuses on simple data that is easy to gather may be seen as requiring less time and expertise to manage. It may also be the case that those interested in accountability see performance measurement as more objective and fair than other forms of evaluation. An accountability system built around a consistent, clearly explained and fixed set of indicators is a system that can make decisions mathematically. The potential for subjective interpretation or bias is reduced.

However, funders sometimes try to stretch performance measurement approaches to serve other purposes. They hope to learn about program impact and they hope to make important, strategic decisions on the basis of this analysis. This expectation may not be realistic.[xlvii]

Furthermore, the idea that performance measurement is simpler, cheaper, and more mechanical than other forms of evaluation may be misguided. When an elegant, well designed performance measurement system is up and running, it may look like nothing more than a set of check-boxes and forms. Yet, good managers, like good pilots, know which indicators to watch because they have a deep understanding of what is going on behind the dashboard. Effective performance measurement systems stand on the shoulders of a great deal of strong communication and clarity of purpose. Buy-in, trust, and clarity of purpose for front-line staff and field managers are arguably even more central to the success of this type of measurement.

There is a mismatch between investments and expectations.

Evaluation work is often not resourced well enough to meet expectations.

Many of the funders who require evaluation do not provide funds for evaluation work.[xlviii] While money is not the only important factor, many nonprofits do not have the capacity to tackle evaluation in a manner that provides them with what they need.[xlix] As well, few nonprofits have neither a trained evaluator on staff nor staff with sufficient knowledge to be able to design and carry out a data collection initiative and make sense of the findings.

In a 2010 study, only 32% of nonprofits felt that foundation funders had been helpful to their ability to measure progress.[l] Meanwhile, staff members at funding organizations often do not have the expertise or the time to respond to questions from their grantees about evaluation, especially if they are smaller foundations.[li] Furthermore, when funders do offer training and support to nonprofits around evaluation, this support often focuses on the process of gathering evaluation data. It is less likely, therefore, to help nonprofits participate as partners in developing the evaluation agenda or to build capacity around synthesizing what they have learned and communicating these insights effectively to various audiences. While it may not be realistic in the short term to expect increased investment in evaluation, it may be possible to make sure that expectations align in a reasonable way with the current level of investment.

Communication is inadequate.

Evaluation becomes more effective when there is ongoing communication across key stakeholders, especially when the evaluation is complex. Evaluation work becomes problematic when it is decoupled from investments in communication.

We know from the research literature that investment in ongoing communication is one of the best ways to ensure that evaluations get used. However, in a 2010 survey, nonprofit leaders from across the United States felt a need for more discussion with their funders about their evaluation work.[lii] In particular, 58% wanted more discussion around interpretation of the results and 71% wanted more discussion around developing the skills to collect and interpret data.[liii]

Most often, in accountability relationships, the flow of information tends to be from the nonprofit to the funder (i.e., the nonprofit reports to the funder on its progress). While this may satisfy the accountability component of an evaluation, it does not necessarily lead to learning or the strengthening of relationships. In other words, when communication is limited to a grant application, a few written progress reports along the way, and a final report at the end, there is little opportunity to address potential concerns that may arise. These concerns, from a nonprofit perspective, can include questions such as how a report it produces will be used by funders, what happens to the information and data that it is rolling up to funders, as well as how to get feedback from the funders on what worked well and what did not.

Imagine a community in which there are five agencies that do work with youth at risk. Let us assume that each of these agencies receives funding for this work from two or three different funders. That means that there may be as many as 15 reports submitted to funders each year, all focused on more or less the same cluster of outcomes.

However, this large cluster of data and insights is never considered in a holistic way. The five agencies are not likely to see one another's reports. The funders do not typically share the reports they receive with each other or provide feedback to the grant recipients. Chances are that the youth served by these programs or the community members who support them through taxes or donations will see none of this information.

Evaluation work is not addressing the full spectrum of questions needed to lead to action.

Effective evaluation answers questions that matter to those who are in a position to take action. Accountability-focused evaluation asks a narrow range of questions.

Undeniably, evaluation is an important part of the nonprofit-funder relationship in helping funders to understand how their dollars are spent and what the effect has been. Yet, while the ideal relationship is one of a partnership with shared responsibility and understanding of each other's needs, in effect the current agenda is often determined by funders. In one study, 53% of nonprofit leaders felt that funders were "primarily interested in information about my organization's performance that will be useful to them, rather than information that provides utility to me and my organization."[liv]

As such, the questions they ask typically address systems-level questions (See Table 3 below) that are less useful for nonprofits.[lv] Sometimes the questions focus on narrow outputs (i.e., how many people attended an event), while other times they attempt to address questions that would be more appropriately addressed through applied research — questions about the interrelationships among multiple interventions, or questions about long-term impacts that are not immediately apparent by the time a nonprofit has to report back to its funder.[lvi]

If local nonprofit groups have little voice in setting Ontario's evaluation agenda, the users of services and community members often have even less. There are few occasions, if any, for community leaders to contribute their input and reflect on the process.[lvii] Furthermore, it can be difficult to build trust when technical evaluation language is used in place of language that is used at the community or service user level.

While nonprofits often feel over-evaluated, the same level of scrutiny is not always applied to other stakeholders in the field. Data in the Ontario context is hard to find, however, a 2014 American survey found that 53% of grantmakers ask their grantees for feedback on their grantmaking work, while 63% sought input on their overall strategy from grantees or community members.[lviii]

Some kinds of evaluation questions — and, in particular, those that are more important to local agencies or to service users — get missed. Accountability-focused evaluation work mandated by funders is not designed to answer the questions that matter to individual nonprofits — questions of personal impact, nuances of local context, and adaptability to changing circumstances. In fact, as Table 3 below suggests, the questions that are of interest to large systems are sometimes in direct conflict with those that are meaningful to individual programs.

Table 3. Competing evaluation questions.

Systems-level evaluation questions	Local evaluation questions
To what degree is this intervention implemented consistently across communities? How can we improve consistency?	How can we adapt this intervention to meet the local needs of our community?
Is this intervention the best use of resources? Should these resources be redirected to more effective measures?	How can we continue to adapt and shape this intervention so that it becomes more efficient and has more impact over time?
What distinct and unique contribution did our investment make?	How did the program pull in different kinds of resources in creative and integrative ways?

There is nothing inherently wrong with an evaluation project that addresses questions important to governments and other funders. These questions are just as valid as those posed by any other stakeholder group. In fact, an argument could be made that government is well positioned to raise evaluation questions that speak to the greater common good because it is accountable to the entire electorate. However, an evaluation agenda driven by a single stakeholder group is not likely to generate good insights for action.

Reflections

Nonprofits often express frustration when it comes to the topic of evaluation. This section has made an effort to frame this problem more clearly. When evaluation is a top-down exercise it can result in the potential for nonprofit organizations to assume "cynical compliance" and "secret resistance" attitudes, whereby "people carry on working according to their own professional judgement, while still reporting up the system what they perceive to be ridiculous numbers."[lix] That is to say that in situations like these, evaluation is seen to be something for someone else rather than a collaborative approach to learning. Rosalind Eyben refers to the influence of evaluation methodologies and tools (she uses the term *artefacts*) as "technologies of power" that are "enforced by authority but internalised so that no obvious external control is required."[lx] In other words, evaluation becomes about a methodology or tool rather than a learning process.

In many respects, the crux of the problem exists at the intersection of our beliefs about learning, action, and accountability. Fundamentally, evaluation is a waste of resources unless it leads to better services that achieve greater change on issues that matter to Ontarians. The evaluation literature is rich with ideas about how to make sure evaluation leads to action. In particular, the literature tells us that action occurs when the stakeholders who are well positioned to act are deeply engaged in the evaluation process. Although there are numerous examples of good utilization-focused evaluation in Ontario, there is also a great deal of evaluation work that does not apply these principles.

What is the problem?

Although much good evaluation work happens in Ontario, nonprofits are most often frustrated with a particular type: externally-mandated evaluation that is undertaken for the purposes of accountability or systems-level evaluation. In large part, this frustration arises because this type of evaluation is not seen to lead to action. Often, the expectations are not a good match with the methodological approaches being used or the level of resources invested. The range of evaluation questions being addressed is too narrow to be useful. There is too much emphasis on the mechanics of data collection and insufficient investment in building buy-in, forming consensus, and articulating intended uses. There is insufficient communication throughout the process, leaving nonprofits unsure about whether the information was used or how. In many respects, these evaluation-related problems are a reflection of larger challenges in the nonprofit sector that centre on how investment is structured and how accountability is conceptualized.

In the following section, we explore some principles that lead to greater utilization of evaluations.

4. Promoting usefulness

Although problematic evaluation happens frequently in the nonprofit sector, it does not happen because of a lack of ideas about how to do better. Front-line nonprofits, funders, evaluators and policy analysts across the province are implementing strategies that are setting the stage for utilization. This section reviews some of these ideas.

When evaluation work is initiated by an external funder, and the focus is only on accountability, the chances that an evaluation will get used are reduced. Individual nonprofits can (and regularly do) overcome these odds and produce useful evaluation work that meets the expectations of their funders. However, they should not have to. There are a number of ideas being explored in various contexts that attempt to make utilization the norm.

Promoting collective impact and strategic philanthropy

Recently, many funders have adopted approaches to grantmaking that are informed by theories like strategic philanthropy and collective impact.[lxi] These approaches to investment are (in part) an effort to create more meaningful, evidence-based accountability procedures that are more firmly tied to clear impact goals and grounded in more participatory partnerships with nonprofit groups. In these approaches, grantmakers work with other stakeholders to identify clearly defined shared goals and a detailed theory of change that explains how they intend to achieve these goals through investments. Achievement of outcomes plays a crucial role in soliciting donations as well as in grantmaking. From the very beginning of the application process, it is made clear to potential grant recipients that they will be responsible for reporting on specific metrics related to funder goals. The need for ongoing investment in communication, measurement, thoughtful analysis of data and transparent sharing of findings is acknowledged.

While these approaches have shown promise, they have also been criticized as promoting overly complex and onerous measurement expectations. Although designed in the spirit of partnership and collaboration, they do not always include mechanisms capable of redressing the power imbalance between funders, nonprofits and service users.[lxii] They are not explicitly designed to broaden the voices involved in setting the evaluation agenda or alter the patterns of communication and interaction among stakeholders. Sometimes, the investment priorities that emerge from community consultation can become rigid over time.

Re-thinking accountability relationships

Evaluation is increasingly important to the funder-nonprofit investment relationship as interest in performance-based funding models continues to grow. Governments and private funders are seeking evidence on program effectiveness as a way to channel (always limited) resources to the highest performing organizations. While data and evaluation practices are still underdeveloped in the nonprofit sector, the demand for evidence to guide funding decisions will likely mean new resources must be dedicated to evaluation for this purpose.

The shift toward performance-based funding models presents both a challenge and an opportunity for nonprofits to critically examine their ways of working and to experiment with new approaches. While there are concerns about tying funding to results, such as the risk of “creaming” the best clients or choosing inappropriate targets for the client group, those nonprofits and sub-sectors that can demonstrate, through evaluation, that they are achieving results will be well placed to innovate and attract new funding. Regardless of whether performance-based funding models are appropriate for particular kinds of programs and services, the opportunity should not be missed to redistribute funding away from poor-performing areas toward higher-performing ones by relying more on evaluation in the service of the investment relationship.

These models look at accountability in somewhat new ways, but some authors advocate for more fundamental change in how we approach accountability. For example, Benjamin (2008) argues that it may be possible to improve the usefulness of evaluation in accountability relationships by reviewing the sources of risk in the relationship and the significance of those risks. For example, accountability systems that are designed to maximize shared learning or support a nonprofit to handle unexpected challenges that arise during the grant term may look quite different from systems focused on preventing a grant recipient from wasting money. As Craig Valters notes, “...there is no reason why, for example, programmes could not be held accountable for how much has been learnt over time, how they have adapted to new information and why this adaption has been important for improved development outcomes. This cannot be the case if accountability continues to be conflated with accountancy.”[lxiii]

Promoting evaluation of evaluation

Many resources are devoted to reporting and evaluation, but we know very little about the return on this investment. Funders rarely evaluate themselves directly and rarely check to see whether the evaluation findings reported to them by grant recipients led to action in the intended ways. However, this is changing. The Ontario Trillium Foundation, for example, has recently announced that it will make make data about its grants publicly accessible. In the U.S., the Center for Effective Philanthropy and Philamplify are developing programs and tools focused on getting feedback from grantees to funders.[lxiv]

Promoting use to advocate policy changes

The potential for evaluations to inform policy at both an organization and legislative level is something that should not be ignored. An evaluation can be an empowering force that brings to light the evidence that shows a need for action. This may not be the goal for an organization starting out but, in certain contexts, it can demonstrate that an issue requires either a new policy or a change in policy and ultimately can affect societal change.

Building capacity to negotiate evaluation agreements

Evaluation is a complex field and the stakeholders in Ontario's nonprofit sector do not always have clear or consistent understandings of its jargon, its uses, or its risks. Although nonprofits have access to many resources that can help them develop evaluation plans internally, there are very few resources designed to help nonprofits negotiate meaningful evaluation agendas with their funders or their partners. The Ontario Federation of Indigenous Friendship Centres (OFIFC) has developed a set of principles and guidelines that its members can use when they are invited to participate in applied research or evaluation work.

Simplifying the evaluation process

Some authors have argued that utilization of evaluation could be improved simply by making the process less technical and complex. For example, it has been argued that evaluation can be done by anyone and that rigorous "technically correct" evaluation is not always called for.[lxv] Instead, depending on the context, nonprofits should embrace imperfect but relevant evaluation that is manageable, useful, and does not require significant extra time or money. Results-based accountability is an approach that advocates for frequent dialogue between partners about the change they are making, focused on "turning the curve" of a specific, simple indicator.[lxvi] Richard Harwood has developed approaches to community consultation that focus on valuing "public knowledge" more and de-emphasizing the role of "expert knowledge."[lxvii]

Making evaluation jargon simpler and more consistent has also been suggested. For example, the use of words like *evidence* and *results* can carry different meanings depending on who is using them (i.e., a funder's definition of what constitutes evidence may differ than that of a nonprofit).[lxviii] By designing processes that use language that is easily understandable across stakeholders, conversations can be opened up that emphasize a shared path.

Expanding the methodological toolbox

In some instances, creative approaches to evaluation should also be considered as they can help to facilitate the conversation and learning that more traditional methods might not allow for. In many cases, these methods are focused on using qualitative data in more systematic ways and challenging the assumption that numbers are always the most effective ways to demonstrate outcomes. For example, a project looking at urban high school-based teen pregnancy prevention programs in Greater Boston used a collage and discussion exercise with participating students to help promote dialogue on what constitutes a healthy relationship.[lxix] Furthermore, non-traditional forms of evaluation can also provide some of the best learning opportunities. For instance, in the arts sector, artists are often their own greatest critics and self-assessment can be a valuable part of an evaluation.[lxx]

Ultimately, evaluations that bring together many different stakeholders are more often going to lead to success. Transformative evaluation, empowerment evaluation, and participatory action research are a few examples of methodologies where learning is a central element. In these methodologies, the primary audience is at the community level rather than the nonprofit or funder.[lxxi] This means involving community stakeholders in designing and implementing the evaluation plan as well as in the sharing of the results.[lxxii] The implication is that evaluation is never value free, nor truly objective. Therefore, community stakeholders should also be part of determining the evaluation process to ensure that their values are represented.

As an example of this principle in action, the OFIFC explicitly prioritizes the community level in its research framework booklet:

> Our approach to research is practical, fully recognizing communities as authors of the knowledge that any community-driven inquiry generates. USAI [Utility Self-Voicing, Access, Inter-Relationality] stresses the inherent validity of Indigenous knowledge and positions it within all relationships, fully acknowledging political context…. USAI also responds to evaluation issues, offering a culturally-relevant approach to measures and indicators.[lxxiii]

Evaluations can help to empower communities to push for change through the generation of information and data as well as through facilitating dialogue with different actors.[lxxiv] It may be that there are different opinions and needs — and, indeed, those differing opinions and needs may be legitimate — but an evaluation that begins by leaving out voices can lead to the building of walls when none should exist.

Other ideas

This list is far from complete. Other ideas include:

- making more strategic use of applied research and bringing academic context experts into the discussion in a stronger way;
- promoting greater collaboration across funders on evaluation to create more consistent jargon-free language and expectations;
- exploring the potential for *networked* evaluation practice where nonprofits might work together on multi-site evaluation or applied research work;

- developing strategies to share evaluation approaches, methods and findings more effectively across the sector.
- using infographics and social media to communicate evaluation findings more effectively;
- using new technology (such as database design) to simplify evaluation; and
- making more use of developmental evaluation approaches that are more flexible and dynamic.

Reflections

Making evaluation more useful, in part, involves developing a deeper, more refined way of thinking about what we mean by evaluation and why we want to do it. In this section, the evaluation-related challenges facing the nonprofit sector and the underlying reasons why they exist were described in a more specific way. It also made it clear that the sector is committed to doing evaluation that gets used and that it has access to a wide range of tools and strategies that have excellent potential to help.

What are some initial ideas that could inform action?

Some funders are setting the stage for more useful evaluation findings by revisiting the relationship between measurement and accountability. They are working to deepen their partnership with grant recipients and community members by developing shared agendas for change. They are clarifying their evaluation expectations and making more explicit commitments to act on what they learn. Funders, nonprofits, umbrella groups and researchers are beginning to expand our notions of how evaluation findings can and should be used and beginning to hold one another more accountable for that kind of use. This type of work is most successful when all parties commit to ongoing communication and participatory decision-making. That work takes practice and training.

These kinds of stage-setting actions may have the effect of integrating evaluation practice into nonprofit management in a more meaningful way. They may also open up the field of front-line program evaluation to a wider range of approaches. Simpler methods that emphasize community engagement and short feedback cycles also have great potential, as do more participatory approaches.

5. An emerging vision for a Sector Driven Evaluation Strategy

Ontario’s nonprofit sector conducts a great deal of evaluation work, and this practice is becoming more and more diverse and innovative over time. Even so, many observers conclude that the agenda for evaluation in Ontario is still set primarily by governments and other funding bodies. The questions being asked are often narrowly focused on accountability and this reduces the potential for meaningful learning and action. The implications of this power imbalance include wasted resources, frustration, misperceptions about the function and use of evaluation, damaged inter-stakeholder relationships and important unanswered questions about impact.

ONN's Sector Driven Evaluation Strategy is meant to resonate with nonprofit organizations across Ontario regardless of mission or size. Consequently, it will be important for the strategy to clearly distinguish the most appropriate uses of performance measurement, program evaluation, systems evaluation and applied research. At the same time, it may also be intended to help integrate and apply the lessons learned through many different evaluation projects undertaken by different people for different reasons. Sometimes, these reasons may be at odds with one another. Even as it attempts to create an integrative vision for evaluation work in the sector, ONN's strategy needs to leave space for different kinds of evaluations, with competing agendas, to co-exist. Perhaps even more crucially, the strategy needs to ensure that service users and community members are involved in setting the evaluation agenda.

The themes explored in this literature review suggest a number of elements that could be included in the Sector Driven Evaluation Strategy.

For instance the strategy could:

1. Raise awareness about the qualities that make an evaluation project useful by:
    a. Underscoring the importance of developing questions in a participatory way, identifying intended users early in the planning process, and offering practical tips for reaching consensus on evaluation questions.
    b. Itemizing the types of questions that make evaluations useful to various types of users.
    c. Helping nonprofits identify situations where applied research might be more useful than evaluation.
    d. Sharing best practices of good process that lead to really useful evaluation.
2. Build capacity at a nonprofit level to negotiate evaluation with different stakeholders.
3. Elevate discussion about the purposes of evaluation work and help to identify evaluation initiatives that lack focus or attempt to serve too many agendas at once. It could flag cases where the balance of evaluation resources should be shifted away from data gathering and toward building stronger collaborative relationships. It could call for an evaluation of evaluations to track usefulness

over time. It could help nonprofits to diagnose situations where an evaluation project is based on unrealistic or misaligned expectations.

4. Clarify roles in Ontario's nonprofit evaluation ecosystem. It may be, for example, that evaluation work would be more useful if it was disentangled from the accountability and reporting process and managed in a more independent way. Independent, multi-sectoral planning tables or provincial networks may be well positioned to facilitate discussions about evaluation priorities because they have no direct power over any of the stakeholders involved.
5. Promote or start conversations among and across different levels of stakeholders (funder-funder, funder-grantee, and grantee-grantee) about their evaluation needs and where there can be alignment.

Final thoughts

This report serves as a first step toward developing the Sector Driven Evaluation Strategy. Many different kinds of measurement work are undertaken in the sector. One simple but important way of setting the stage for evaluation use is to become more informed about the relative strengths and limitations of applied research, performance measurement, program evaluation and systems evaluation. This will help us to more quickly diagnose situations where expectations and strategies are misaligned. There are other important changes in perspective that have the potential to be very helpful. Re-orienting evaluation practice to put engagement and relationship building at the centre of the process is a powerful step.

In Ontario's nonprofit sector, evaluation is least likely to lead to positive action when it is used only as a means to hold nonprofits accountable for their use of grant money. The potential for learning and action is even lower if the process is poorly explained, based on unrealistic expectations, or under resourced. This is not in the least surprising. The evaluation literature has identified these mistakes and their implications time and again. It has also developed a wide range of ideas about why these problems occur and how to avoid them. It may be fair to say that the key problem with evaluation in the nonprofit sector is the *misuse* of evaluation. When evaluation is misused, the people involved can be treated like the raw materials in a mechanical process rather than as partners with power and voice. Although this paper has focused on the ways in which evaluation fails to produce useful results, it is also important to say that evaluation can cause serious harm. Misuse of evaluation can be a way to avoid engaging in difficult conversations with partners. It can undermine privacy and dignity for individuals and it can be used to rationalize poor policy decisions.

Talking about evaluation as a dry, technical undertaking has led us, perhaps, to understate its importance. At the same time, the investment in measurement and evaluation work has been huge. Yet, although the data on the impact of evaluation is not sufficient to reach strong conclusions, the ideas reviewed in this paper strongly suggest that the sector's investment in evaluation to date has not always been as informed or grounded as it should be. Consequently, it has not yielded as much value as it should and it has sometimes had negative consequences.

Going forward, we are beginning to envision what a strategy for evaluation in the nonprofit sector could look like. Yet, we also know our limitations. There are other reports, other people, and other issues that we may have missed and as such we do not claim that this paper is an exhaustive summation but rather acts as a snapshot of what we have read, heard, and seen to date. For this reason, we recognize that it is only through further engagement with the sector that we will be able to move past some of the issues we've raised here and begin to shift the perspective about what evaluation can and should be.

In that sense, for us to achieve our goals, we will need your help and we invite you to get in touch and provide us with your feedback. Help us develop the Sector Driven Evaluation Strategy and make evaluation less about paperwork and more about insight. The stakes are higher than you might think.

We want to hear from you!

Give us your feedback on this report: https://theonn.wufoo.com/forms/evaluation-literature-review-feedback/

For more information, visit: http://theonn.ca/our-work/our-structures/evaluation/

Stay connected to our work: Subscribe to the Evidence Network News Flash

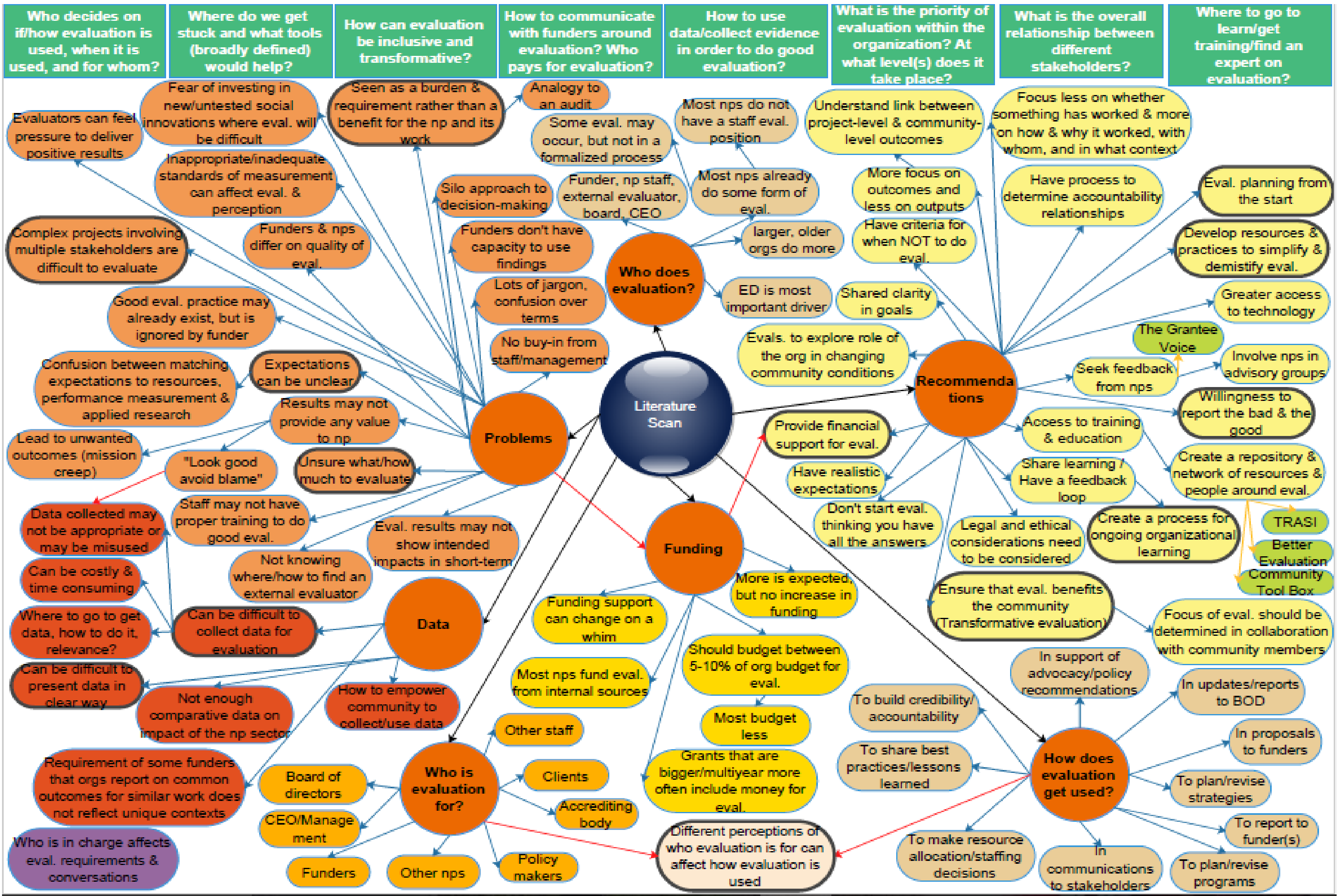
Glossary

Accountability	"A process of holding actors responsible for actions" (Fox and Brown 1998, 12).
Applied research	As compared to evaluation, applied research is often more time consuming, more theory-driven, and more expensive. It often involves consideration of data from multiple programs or program sites. It is typically designed and carried out by academic researchers who are content experts in the research topic under consideration.
Culture of learning	Whereby nonprofits use evaluation as a way to reflect, learn, and share their failures and successes (as opposed to using evaluation as simply an accountability exercise).
Evaluation	"Evaluation is the systematic collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding (Patton 2008, 39).
Look-Good-Avoid-Blame (LGAB)	A mindset whereby people attempt to minimize failure so as not to be seen as responsible for it.
New Public Management (NPM)	A governance strategy that began in the late 1970s and "includes linking resource allocation to performance, competition between providers of services, greater discipline and parsimony in resource use, and adoption of what is represented as private-sector management practices" (Eyben 2013, 13).
Performance management	The ongoing, day-to-day data gathering that program staff and volunteers do as part of their job. This type of measurement is typically integrated into the day-to-day work of managing a program.
Program evaluation	Involves deeper investigation about the outcomes or impacts as well as questions about process. It attempts to determine whether or not a program led to change, and why.
Sector Driven Evaluation Strategy	ONN's project to design a strategy to engage the sector on the systemic issues around evaluation for Ontario's nonprofit sector and offer up some principles and solutions to move forward

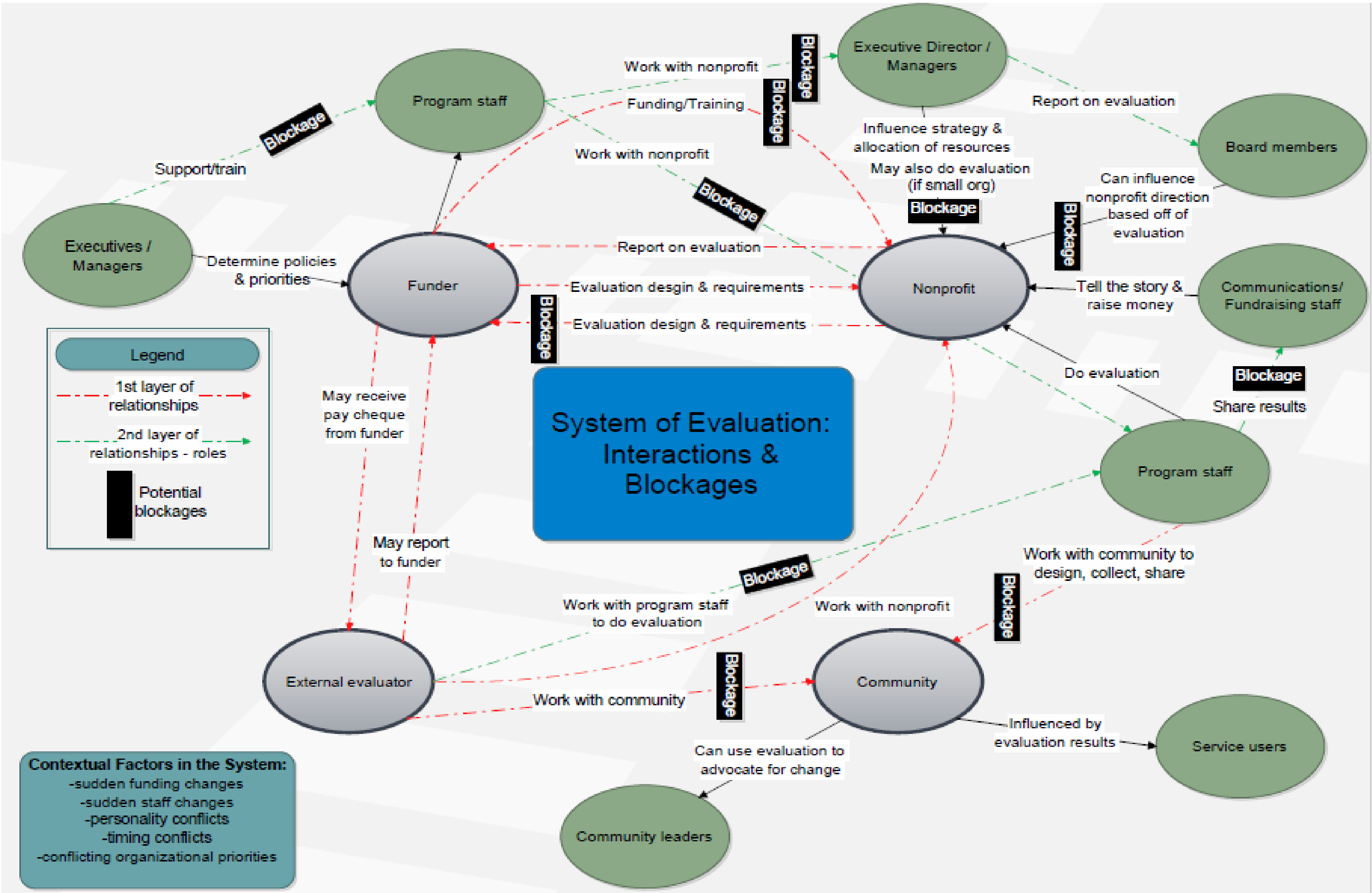
	based on sector feedback.
Systems change	"Systems change is an intentional process designed to alter the status quo by shifting the function or structure of an identified system with purposeful interventions. It is a journey which can require a radical change in people's attitudes as well as in the ways people work. Systems change aims to bring about lasting change by altering underlying structures and supporting mechanisms which make the system operate in a particular way. These can include policies, routines, relationships, resources, power structures and values" (Abercrombie, Harries, and Wharton 2015, 9).
Systems evaluation	Evaluation work focused on the study of complex and interconnected sets of interventions.
Theory of change	The set of assumptions that explain both the mini-steps that lead to the long-term goal of interest and the connections between program activities and outcomes that occur at each step of the way.
Utilization-Focused Evaluation (UFE)	An approach, designed by Michael Quinn Patton, that tries to maximize the potential for action by emphasizing the importance of identifying the intended uses and users of an evaluation in the planning stages.
- Audience	All those who might have a passive interest in reading a report
- Evaluation *uses*	
- Conceptual use	Occurs when an evaluation helps users develop a deeper understanding of key issues or ideas.
- Instrumental use	Occurs when an evaluation's recommendations are acted upon directly.
- Process use	Occurs when the evaluation experience itself leads to action.
- Symbolic, legitimative, or persuasive use	Occurs when an evaluation's findings help to bolster the case for an intervention that is already planned, or to adjust the strategy for implementation of that intervention.
- Intended users	Those who are expected to take specific types of action on the basis of the evaluation.

Voluntary Sector Initiative	Initiative by the Canadian federal government and the nonprofit sector in the early 2000s to improve nonprofit capacity and the relationship between the federal government and the sector.

Appendix A: Literature review mind map



Appendix B: Evaluation system map



Appendix C: The differences between tools

Table 4. Differences between tools.

	Performance Measurement	**Program Evaluation**	**Systems Evaluation**	**Applied Research**
Purpose	→ongoing monitoring →minor program adjustments →accountability	→learning →consensus building →critical review demonstration of impact of specific programs in specific contexts (often short-term) →identification of specific, localized recommendations for future action by specific actors	→demonstration of the impact of a network of connected interventions →identification of high level recommendations for future action	→generation of new, generalizable knowledge about best practices →study of the typical long-term impact of program types or models
Methods	→simple, non-intrusive methods like intake forms, checklists, and simple feedback surveys. →often focused on processes and outputs →usually highly structured and	→wider range of more intensive qualitative and quantitative methods →more use of pre-post designs →more adaptation of methodologies as the project unfolds	→aggregation of data from many sources	→very wide range of methods →longitudinal approaches, quasi-experimental designs

	quantitative →static through time			
Roles	→data collection conducted by staff →analysis by manager on an ongoing basis →IT or database specialists may be involved	→staff and management involved →may also involve specially trained internal or external evaluators →often guided by cross-stakeholder committees		→often led by academics or other content area specialists →often guided by cross-stakeholder committees

Table 5. Purposes and approaches.

Purposes or expectations	Approaches			
	Performance Measurement	**Program Evaluation**	**Systems Evaluation**	**Applied Research (New knowledge)**
Programs, sites or agencies will be held accountable for delivering the agreed upon work	Good fit! ... if data is used!	Not a good fit. →Evaluation methods are more complex than needed.	Not a good fit.	Not a good fit.
Programs, sites or agencies will be held accountable for the impact they achieve	Can work well... →If performance measurement systems are sophisticated, specialized, and carefully monitored.	Can work well... →But evaluations undertaken for this purpose may not be as good at generating local insights or actions (see below).	Can work well... →When time and energy are invested in shared measurement systems, ongoing communication, backbone infrastructure, and a shared sense of purpose.	Not a good fit. →Applied research is not designed to inform action in a direct way.
Local programs or agencies will develop insights about their work and its impact leading them to improve practice	Rarely works. →Measures are focused on outputs, buy-in is minimal and analysis is basic.	Good fit! →Especially when time and energy is invested in buy-in, communication, clarity of purpose, and plans for use.	Can work well... →If local sites are engaged as partners in the process.	
The sector or the community will	Rarely works on its own.	Rarely works. →Local evaluations do not typically	Good fit! →Especially when time and energy are invested	Good fit! →Especially when the

develop new knowledge about best practices and long-term impacts	→Although performance measurement methods are often useful when incorporated into systems evaluation projects.	measure long term change and are not designed to generate generalizable knowledge.	in shared measurement systems, ongoing communication, backbone infrastructure, and a shared sense of purpose.	research questions are highly focused and specialized.
Large systems will develop insights about their work and its impact leading them to improve practice		Rarely works. →It is challenging to aggregate findings from disparate local evaluations.		Can work well... →If research findings are presented in an accessible way and provided in a timely manner.

Works cited

Abercrombie, Harries, and Wharton. 2015. *Systems Change: A guide to what it is and how to do it.* Report, LankellyChase Foundation and NPC.

Alaimo, Salvatore P. 2008. "Nonprofits and Evaluation: Managing Expectations from the Leader's Perspective." Edited by J.G. Carman and K.A. Fredericks. *New Directions for Evaluation* (Nonprofits and evaluation) 73-92.

Ball, Ted. 2014. *HOW TO STOP BLAMING: Six Principles For Accountability Design.* January 13. Accessed August 6, 2015. http://quantumtransformationtechnologies.com/wp content/uploads/2014/01/How-To-Stop-Blaming-Six-Principles-Of-Accountability-Jan-13 2014.pdf.

Benjamin, L. 2008. Evaluator's role in accountability relationships: Measurement technician, capacity building or risk manager? *Evaluation*, 14 (3), 323-343.

Behrens, Teresa R., and Thomas Kelly. 2008. "Paying the Piper: Foundation Evaluation Capacity Calls the Tune." Edited by J.G. Carman and K.A. Fredericks. *New directions for evaluation* (Nonprofits and evaluation) 37-50.

Brock, Andrea, Ellie Buteau, and An-Li Herring. 2012. *Room for Improvement: Foundations' Support of Nonprofit Performance Assessment.* Report, Cambridge: The Center for Effective Philanthropy. Accessed August 4, 2015. http://www.effectivephilanthropy.org/wp content/uploads/2014/-01/RoomForImprovement.pdf.

Brock, Kathy L. 2010. "A Comprehensive Canadian Approach to the Third Sector: Creative Tensions and Unexpected Outcomes." In *Policy Initiatives Towards the Third Sector in INternational Perspective*, edited by Benjamin Gidron and Michal Bar. New York: Springer New York.

Brodhead, Tim. 2010. *On Not Letting a Crisis Go to Waste: An Innovation Agenda for Canada's Community Sector.* Report, The J.W. McConnell Family Foundation and Social Innovation Generation.

Campbell, David. 2002. "Outcomes Assessment and the Paradox of Nonprofit Accountability." Nonprofit Management & Leadership 243-259.

Carman, J.G., and K.A. Fredericks. 2008. "Nonprofits and Evaluation: Empirical Evidence from the Field." Edited by J.G. Carman and K.A. Fredericks. *New Directions for Evaluation* (Nonprofits and evaluation) 51-71.

Carman, Joanne G. 2008. "The Accountability Movement: What's Wrong With This Theory of Change?" *Nonprofit and Voluntary Sector Quarterly* 256-274.

Carpineto, Julie and Kelly Washburn. 2014. *Julie Carpineto and Kelly Washburn on Using a Creative Evaluation Project to Engage Youth.* June 13. Accessed September 22, 2015. http://aea365.org/blog/julie-carpineto-and-kelly-washburn-on-using-a-creative-evaluation-project-to-engage-youth/.

Center for Effective Philanthropy (CEP). 2014. “Grantee Perception Report (GPR) and Declined Applicant Perception Report (APR).” Accessed November 12, 2015. http://www.effectivephilanthropy.org/assessments/gpr-apr/.

Connell, J., and A. Kubisch. 1998. "Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects and Problems." In *New Approaches to Evaluating Community Initiatives, vol. 2, Theory, Measurement, and Analysis*, edited by K. Fulbright-Anderson, A. Kubisch and J. Connell. Washington, DC: Aspen Institute.

David, Tom. 2006. *Evaluation and Foundations: Can We Have an Honest Conversation?* March 21. Accessed July 27, 2015. http://nonprofitquarterly.org/2006/03/21/evaluation-and foundations-can-we-have-an-honest-conversation/.

Dyson, Alan and Liz Todd. 2010. "Dealing with complexity: theory of change evaluation and the full service extended schools initiative." *International Journal of Research & Method in Education 33* (2), 119-134.

Eyben, R. 2013. *Uncovering the Politics of 'Evidence' and 'Results'. A Framing Paper for Development Practitioners.* Big Push Forward.

Feller, I. 2002. Performance measurement redux. *American Journal of Evaluation*, 23 (4), 435-452.

Fox, J. A. and L. D. Brown (eds). 1998. *The Struggle for Accountability: The World Bank, NGOs, and Grassroots Movements,* Cambridge, MA: The MIT Press.

Fram, Eugene H., and Jerry L. Talley. 2012. *Using Imperfect Metrics Well: Tracking Progress and Driving Change.* July 24. Accessed August 5, 2015. http://nonprofitquarterly.org/2012/07/24/using-imperfect-metrics-well-tracking-progress-and-driving-change/.

Friedman, M. 2005. *Trying Hard is Not Good Enough.* Author.

Grantmakers for Effective Organizations (GEO). 2015. "How Can We Help Our Grantees Strengthen Their Capacity for Evaluation?" Accessed August 13, 2015. http://docs.geofunders.org/?filename=evaluation_capacity_building.pdf.

Grantmakers for Effective Organizations (GEO). 2015. *Learning Together: Actionable Approaches for Grantmakers.* Report, Washington.

Grantmakers for Effective Organizations (GEO). 2014. *Is Grantmaking Getting Smarter?* Report, Washington.

Grantmakers for Effective Organizations (GEO). 2011. “How Do We Build the Capacity of Nonprofits to Evaluate, Learn and Improve?” Briefing Paper, Washington: Grantmakers for Effective Organizations. Accessed August 4, 2015. http://docs.geofunders.org/?filename=GEO_SWW_BuildCapacityToEvaluateLearnImprove.pdf.

Hall, Michael H., Susan D. Phillips, Claudia Meillat, and Donna Pickering. 2003. *Assessing Performance: Evaluation Practices & Perspectives in Canada's Voluntary Sector.* Research Report, Ottawa: Canadian Centre for Philanthropy and Centre for Voluntary Sector Research and Development.

Harwood, R. 2005. “Standards of Excellence in Civic Engagement: How Public Agencies Can Learn from the Community, Use What They Learn, and Demonstrate that Public Knowledge Matters.” The Harwood Institute. Accessed November 12, 2015. http://c.ymcdn.com/sites/www.iap2.org/resource/resmgr/imported/ResourceMaterials_StandardsofExcellence.pdf

Hogan, Paul. 2013. *Thinking About Nonprofit Evaluation as Affected by Time.* September 4. Accessed July 27, 2015. http://nonprofitquarterly.org/2013/09/04/thinking-about-nonprofit-evaluation-as-affected-by-time/.

Holley, Marc J., and Matthew J. Carr. 2014. *The Role of Evaluation in Strategic Philanthropy.* March 7. Accessed July 27, 2015. http://nonprofitquarterly.org/2014/03/07/the-role-of-evaluation-in-strategic-philanthropy/.

Hoole, Emily, and Tracy E. Patterson. 2008. "Voices From the Field: Evaluation as Part of a Learning Culture." Edited by J.G. Carman and K.A. Fredericks. *New Directions for Evaluation* (Nonprofits and Evaluation) 93-113.

Kahan, Barbara. 2008. *Excerpts from Review of Evaluation Frameworks.* Report, Saskatchewan Ministry of Education.

Kania, M., and J. Kramer. 2011. *Collective Impact.* Stanford Social Innovation Review. Leland Stanford Jr. University.

Kosheleva, Natalia. n.d. *Methodologically Inclusive Transformative Evaluation as an Enabler of Social Inclusion.* Background Paper, International Organization for Cooperative Evaluation.

Lee, Kien, and David M. Chavis. 2015. *How Evaluation Can Strengthen Communities.* June 18. Accessed August 19, 2015. http://ssir.org/articles/entry/how_evaluation_can_strengthen_communities.

Liket, Kellie C., Marta Rey-Garcia, and Karen E.H. Maas. 2014. "Why Aren't Evaluations Working and What to Do About It: A Framework for Negotiating Meaningful Evaluation in Nonprofits."*American Journal of Evaluation* 171-188.

Mayhew, Fred. 2011. "Integrating the Funder-Fundee Relationship into a Model of Evaluation Utilization." *Journal of MultiDisciplinary Evaluation* 1-19.

McMaster, Sir Brian. 2008. *Supporting Excellence in the Arts: From Measurement to Judgement.* Report, London: United Kingdom Department for Culture, Media and Sport.

Morariu, Johanna, Katherine Athanasiades, and Ann Emery. 2012. *State of Evaluation 2012: Evaluation Practice and Capacity in the Nonprofit Sector.* Report, Innovation Network.

National Committee for Responsive Philanthropy (NCRP). 2014. “About Philamplify”. Accessed November 12, 2015. http://philamplify.org/about-philamplify/.

Nelson, Geoffrey, Joanna Ochocka, Kara Griffin, and John Lord. 1998. ""Nothing About Me, Without Me": Participatory Action Research with Self-Help/Mutual Aid Organizations for Psychiatric Consumer/Survivors." American Journal of Community Psychology 881-912

Newcomer, K and C. Brass. 2015. “Forging a Strategic and Comprehensive Approach to Evaluation Within Public and Nonprofit Organizations: Integrating Measurement and Analytics Within Evaluation”. *American Journal of Evaluation*, 36 (1), 1-20.

Ontario Federation of Indigenous Friendship Centres. 2012. *USAI Research Framework.* Research Framework, OFIFC.

Patton. 2010. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use.* Guilford Press.

Patton, Michael. 2008. Utilization-focused Evaluation, 4th Edition. Thousand Oaks, CA: Sage.

Pawson, Ray, Trisha Greenhalgh, Gill Harvey, and Kieran Walshe. 2005. "Realist review–a new method of systematic review designed for complex policy interventions." *Journal of health services research & policy* 21-34.

Peters, R. DeV., Nelson, G., Petrunka, K., Pancer, S.M., Loomis, C., Hasford, J., Janzen, R., Armstrong, L., & Van Andel, A. 2010. “Investing in Our Future: Highlights of Better Beginnings, Better Futures Research Findings at Grade 12.” Kingston, ON: Better Beginnings, Better Futures Research Coordination Unit, Queen’s University.

Ramos, Henry A.J. 2015. *Shortcomings of Modern Strategic Philanthropy and How to Overcome Them.* Summer. Accessed August 31, 2015. http://ncrp.org/publications/responsive-pubs/rp-archive/responsive-philanthropy-summer15/shortcomings-of-modern-strategic-philanthropy-and-how-to-overcome-them.

Ranghelli, Lisa, and Yna Moore. 2015. *Are Nonprofits Ready to Give Foundations Direct Feedback.* June 8. Accessed July 27, 2015. https://nonprofitquarterly.org/2015/06/08/are-nonprofits-ready-to-give-foundations-direct-feedback/.

Scriven, M. 1991. *Evaluation Thesaurus (4th Edition).* California: Sage.

Sharkey, Sarah-Kate G. 2005. "An Exploration of Foundation and Nonprofit Partnerships in Empowerment Evaluation." Department of Planning, Public Policy and Management at the University Oregon, June.

Sherriff, Bronwyn, and Stephen Porter. n.d. "An Introduction to Empowerment Evaluation: Teaching Materials." Accessed August 4, 2015. http://www.mrc.ac.za/crime/evaluation.pdf.

Sieppert, Jackie D. 2005. "Evaluation in Canada's Social Services: Progress, Rifts, and Challenges." *The Canadian Journal of Program Evaluation* 101-121.

Tarsilla, Michelle. 2010. "Inclusiveness and Social Justice in Evaluation: Can the Transformative Agenda Really Alter the Status Quo? A Conversation with Donna M. Mertens." *Journal of MultiDisciplinary Evaluation* 102-113.

Treasury Board of Canada. 2002. Case studies on the uses and drivers of effective evaluations in the Government of Canada. Ottawa: Author.

Thomson, Dale E. 2011. "The role of funders in Driving Nonprofit Performance Measurement and Use in Strategic Management." *Public Performance & Management Review* 54-78.

Torjman, Sherri. 2006. "There's Madness to This Method." *Caledon Institute of Social Policy* 1-4.

Twersky, Fay, and Karen Lindblom. 2012. *Evaluation Principles and Practices.* Internal Working Paper, The William and Flora Hewlett Foundation.

Valters, Craig. 2015. *Four principles for Theories of Change in global development.* September 15. Accessed November 4, 2015. http://www.odi.org/comment/9882-four-principles-theories-change-global-development.

Woolford, A., & Curran, A. 2012. Community Positions, Neoliberal Dispositions: Managing Nonprofit Social Services Within the Bureaucratic Field. *Critical Sociology*, 45-63.

Endnotes

[i] Kahan 2008
[ii] Patton 2008, 39
[iii] Morariu, Athanasiades and Emery 2012
[iv] Mayhew 2011
[v] Readers interested in a comprehensive review of evaluation types are referred to http://betterevaluation.org/approaches.
[vi] Newcomer and Brass 2015
[vii] Newcomer and Brass 2015
[viii] Newcomer and Brass 2015
[ix] Peters, et al. 2010
[x] Scriven 1991
[xi] Pawson, et al. 2005
[xii] Pawson, et al. 2005
[xiii] Dyson and Todd 2010
[xiv] Patton 2008
[xv] Recall, from the previous section, that *use* need not be instrumental in order to be meaningful.
[xvi] Patton 2008
[xvii] Patton 2008
[xviii] Fred Mayhew 2011
[xix] Patton 2008; Mayhew 2011
[xx] Mayhew 2011
[xxi] Patton 2008
[xxii] Patton 2008
[xxiii] Treasury Board of Canada 2002
[xxiv] Alaimo 2008
[xxv] GEO 2015; Alaimo 2008
[xxvi] GEO 2015, Hoole and Patterson 2008, Alaimo 2008
[xxvii] Carman and Fredericks 2008
[xxviii] Morariu, Athanasiades, and Emery 2012
[xxix] GEO 2014
[xxx] Behrens and Kelly 2008; Hall, et al. 2005
[xxxi] Sieppert 2005
[xxxii] Eyben 2013; Woolford and Curran 2012
[xxxiii] Eyben 2013, 13
[xxxiv] Woolford and Curran 2012, 55
[xxxv] Brock 2010
[xxxvi] Torjman 2006
[xxxvii] Ball 2014; Brodhead 2010
[xxxviii] Hall, et al. 2003
[xxxix] Brock, Buteau, and Herring 2012
[xl] Fram and Talley 2012
[xli] Ramos 2015
[xlii] Brock, Buteau and Herring 2012; GEO 2011; Behrens and Kelly 2008; Carman 2008; David 2006
[xliii] Torjman 2006
[xliv] Ranghelli and Moore 2015; Holley and Carr 2014; Ball 2014; Campbell 2002
[xlv] Liket, Rey-Garcia and Maas 2014; Mayhew 2011

[xlvi] Tarsilla 2010
[xlvii] Feller 2002
[xlviii] GEO 2015; Behrens and Kelly 2008
[xlix] Sieppert 2005
[l] Brock, Buteau and Herring 2012
[li] GEO 2015; Twersky and Lindblom 2012; David 2006
[lii] Brock, Buteau and Herring 2012
[liii] Brock, Buteau and Herring 2012
[liv] Brock, Buteau and Herring 2012, 9
[lv] Thomson 2011; Carman and Fredericks 2008
[lvi] Hogan 2013
[lvii] Campbell 2002
[lviii] GEO 2014
[lix] Eyben 2013, 13
[lx] Eyben 2013, 8
[lxi] Holley and Carr 2014; Kania and Kramer 2011
[lxii] Holley and Carr 2014
[lxiii] Valters 2015
[lxiv] CEP 2014; NCRP 2014
[lxv] Fram and Talley 2012
[lxvi] Friedman 2005
[lxvii] Hardwood 2005
[lxviii] Eyben 2013
[lxix] Carpineto and Washburn 2014
[lxx] McMaster 2008
[lxxi] Tarsilla 2010; Sharkey 2005; Nelson, et al. 1998; Sherriff and Porter n.d.
[lxxii] Tarsilla 2010; Kosheleva n.d.
[lxxiii] OFIFC 2012, 5
[lxxiv] Lee and Chavis 2015